REPORT RESUMES

ED 003 557 56

EXPERIMENTAL ANALYSES OF THE EFFECTS OF VARIOUS MODES OF ITEM PRESENTATION ON THE SCORES AND FACTORIAL CONTENT OF TESTS ADMINISTERED BY VISUAL AND AUDIOVISUAL MEANS--A PROGRAM OF STUDIES BASIC TO TELEVISION TESTING.

BY- CURTIS, H.A. KROPP, RUSSEL P.
FLORIDA ST. UNIV., TALLAHASEE, SCH. OF EDUCATION

REPORT NUMBER NDEA-VIIA-385 PUB DATE 61
GRANT OEG-7-08-075
EDRS PRICE MF-$0.18 HC-$3.56 89P.

DESCRIPTORS- INTELLIGENCE TESTS, ACHIEVEMENT TESTS, INTEREST TESTS, *MEDIA RESEARCH, *TEST CONSTRUCTION, ITEM ANALYSIS, TALLAHASSEE, FLORIDA

THE POSSIBILITY OF ADMINISTERING TESTS BY PROJECTED METHODS WAS EXPLORED, FOCUSING ON THE ASSIGNMENT OF VALUES TO SPECIFIC TEST VARIABLES IN ORDER TO MAKE THE RESULTS OF PROJECTED TESTS COMPARABLE TO THOSE OF A NORMAL (CLASSROOM TEACHER ADMINISTERED) TEST ENVIRONMENT. THE VARIABLES OF (1) INTRA-ITEM EXPOSURE INTERVAL, (2) INTER-ITEM EXPOSURE INTERVAL, (3) ITEMS PER EXPOSURE, (4) AUDIOVISUAL AND VISUAL ITEM PRESENTATION, AND (5) ITEM-DIFFICULTY ARRANGEMENT WERE INCORPORATED IN THE PROJECT. CONTENTS OF INTELLIGENCE, ACHIEVEMENT, VOCATIONAL INTEREST, AND ADJUSTMENT TESTS WERE SUBJECTED TO STUDY. STILL SLIDES WERE CHOSEN FOR PROJECTING THE TEST DATA. THESE SLIDES WERE EASILY AND INEXPENSIVELY PRODUCED AND WERE ADAPTABLE TO VARIED ADMINISTRATIVE CONDITIONS FOR TESTING. DURING TESTING, SOME OF THE TEST ITEMS WERE READ FROM TAPE AS THEY APPEARED ON A SCREEN. SUBJECTS RANGED FROM ELEMENTARY TO GRADUATE COLLEGE STUDENTS, DEPENDING ON THE LEVEL OF DIFFICULTY OF A PARTICULAR TEST ADMINISTERED. RESULTS DEMONSTRATED SOME AMOUNT OF EQUIVALENCE BETWEEN THE PROJECTED TEST MODE AND NORMAL TESTING CONDITIONS, ALTHOUGH PAIRS OF MEANS AND VARIANCES WERE AT TIMES SIGNIFICANTLY DIFFERENT. ONE METHOD WOULD BE REGARDED AS INTERCHANGEABLE WITH THE OTHER WHEN THE CORRELATION BETWEEN THE SETS OF SCORES OBTAINED UNDER THE TWO ADMINISTRATIVE CONDITIONS WAS HIGH (APPROACHING OR EQUAL TO THE RELIABILITY OF THE TEST WHEN DETERMINED ON A TEST-RETEST BASIS UNDER NORMAL CONDITIONS). (JH)

ED003557

385

EXPERIMENTAL ANALYSES OF THE EFFECTS OF VARIOUS MODES OF ITEM PRESENTATION ON THE SCORES AND FACTORIAL CONTENT OF TESTS ADMINISTERED BY VISUAL AND AUDIO-VISUAL MEANS: A Program of Studies Basic To Television Testing.

H.A. Curtis and Russell P. Kropp

Title VII Project Number 385

National Defense Education Act of 1958

Grant Number 7-08-075

Department of Educational Research and Testing

School of Education Florida State University Tallahassee

The Research Reported Herein Was

Supported by a Grant from the

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

EXPERIMENTAL ANALYSES OF THE EFFECTS OF VARIOUS MODES OF ITEM PRESENTATION ON THE SCORES AND FACTORIAL CONTENT OF TESTS ADMINISTERED BY VISUAL AND AUDIO-VISUAL MEANS:

A Program of Studies Basic To Television Teaching

H.A. Curtis and Russell P. Kropp

Title VII Project Number 385
National Defense Education Act of 1958
Grant Number 7-08-075

Department of Educational Research and Testing
School of Education Florida State University Tallahassee

The Research Reported Herein Was
Supported by a Grant from the
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education

PARTICIPANTS*

PRINCIPAL INVESTIGATORS

H.A. Curtis

Russell P. Kropp

PARTICIPATING FACULTY MEMBERS

R.O. Billett

F.J. King

Howard W. Stoker

GRADUATE RESEARCH ASSISTANTS

W.L. Bashaw

William L. Burr

Jacob G. Beard

William R. Crawford

Isaiah S. Moyel

Mary Esther Raker

Robert N. Willis

---

*At one time or another during the period of the grant all participants were members of the Department of Educational Research and Testing, School of Education, Florida State University.

ACKNOWLEDGMENTS

The principal investigators gratefully acknowledge their indebtedness to the grantors and publishers who directly and indirectly made possible the research reported herein and to the many individuals whose interest, assistance, and counsel contributed much to the actual conduct of the work itself.

The research reported herein was supported in major part under provisions of Title VII of the National Defense Education Act of 1958 --Grant Number 7-08-075., Project Number, 385. Preliminary studies were supported by a grant to the investigators by the Research Council of the Florida State University. Statistical analyses of the data reported herein were partially subsidized by a grant to the Florida State University from the National Science Foundation, NSF Grant No. G17467.

The investigators are very appreciative of the granting by each of the following publishers of permission to reproduce one or more of their respective tests in experimental form: California Test Bureau, Educational Testing Service, Harcourt, Brace and World, Inc., Psychological Corporation, and Science Research Associates.

Grateful appreciation is also expressed to the following people and agencies for their cooperation in providing data and other forms of assistance throughout the project: fellow faculty members of the Florida State University; the administrative staff of the Florida School for Boys at Marianna; and the staffs of the following Florida county school systems--Bay, Escambia, Hillsborough, Lake, Madison, Pinellas, and Washington.

The investigators are much indebted to Dr. Nicholas A. Fattu, Director, Institute of Educational Research, Indiana University, for the stimulating and sound suggestions which he offered as research consultant during the course of these studies.

While the investigators gratefully acknowledge the contributions of the many organizations and people whose support and assistance made this study possible, they assume full responsibility for the conduct of the research and for the report submitted herewith.

H.A. Curtis

Russell P. Kropp

CONTENTS

SECTION PAGE

I. INTRODUCTION 1

II. PROCEDURES AND APPARATUS 3

III. INTELLIGENCE 6

School and College Ability Test 6

Terman Concept Mastery 18

Terman Concept Mastery: Threshold Times 22

Primary Mental Abilities 24

IV. ACHIEVEMENT 30

Social Studies 35

Reading 37

Arithmetic: I 41

Arithmetic: II 45

V. INTEREST 49

VI. ADJUSTMENT AND PERSONALITY 58

Delinquency-Proneness: I 58

Delinquency-Proneness: II 61

Minnesota Multiphasic Personality Inventory 64

VII. RESPONSE-MARKING TIME 73

Response-Marking Time: Two-Choice Items 73

Response-Marking Time: Four-Choice Items 75

VIII. CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS 76

I. INTRODUCTION

A series of studies was conducted to explore the possibility of administering group standardized tests by projected methods. "Administering," in the sense used here, is intended to encompass all activities ranging from giving initial directions, to presenting items, to giving posttest directions. "By projected methods," is intended to convey that all test items, directions, etc., were presented visually by projecting them on a viewing screen. Essentially, the series of studies dealt with the possibility of substituting the projected presentation of test materials for the traditional test booklet presentation of test materials.

The practical concern underlying these studies is rather obvious: Can standardized, group tests, of the kind commonly used in public schools, be administered by television? A less obvious practical concern was: Would it be possible to program a teaching-testing machine to serve as an "automatic psychometrist." (We believe that the basic variables involved in television testing would be identical with those which would be encountered in programming a teaching-testing machine for test administration.) Thus, our ambition was to explore the possibility of truly mass testing by utilizing educational television networks as a medium by which to present test materials to students. We were mindful that such a procedure might lead to economies in capital outlay for testing materials and costs incurred in shipping test materials. We were aware that the uniformity of test administrative conditions would be a further gain.

Theoretical concerns which precipitated the series of studies were (a) whether projected testing would enable making behavioral appraisals not permitted by paper-and-pencil testing and (b) whether it would permit altering presently available standardized tests so that they might be administered more efficiently, e.g. less test-taking time, without altering the validity of the obtained scores. The studies emphasized less heavily the theoretical concerns than they did the chief practical concern, whether this kind of test administration has any merit at all.

The studies reported herein are connected by an inner logic or strategy that perhaps will remain forever a secret to all except the investigators. Describing it would be tedious, protracted, and, probably, unconvincing. It is sufficient to write that our strategy was to adapt tests, bearing on each of the typical school uses, for projected administration, to administer them, to analyze the data, to compare the results with results obtained under usual test administrative conditions, then to re-adapt, etc. The general goal was broad-scale exploration. The specific goal which dictated each "next step" in the broad attack was the identification of the pertinent manipulatable variables underlying the successful projected administration of tests.

In this general way the following test contents were explored: intelligence, achievement, vocational interest, and adjustment. The following variables, which were known in advance or were isolated en route, were incorporated in most all studies; intra-item exposure interval, inter-item exposure interval, items per exposure, audio-visual and visual item presentation, and arrangement of item difficulty. Values assigned some of the above variables were determined by considering still other variables, e.g., intra-item exposure interval was determined by

(a) dividing the total test taking time, prescribed for the normal administration, by the number of items; (b) by allocating time according to previously determined item difficulty; and (c) allocating time according to the reading level characterizing the item.

Subsequent sections of this report are entitled according to areas of testing--intelligence, achievement, interest, and adjustment and personality--which characterize public school usage of tests. An additional section entitled "Response-Marking Time" contains two studies dealing with the establishment of minimum durations for intra-item exposure intervals. Also, one section is devoted exclusively to the procedures and apparatus used in the development of and the adaptation of testing materials for projected administration. Although we would prefer to have sections of this report arranged according to basic variables, it was not feasible to do so because the majority of variables appeared in each study and the interaction of variables, in some cases, precludes a sensible discussion of one variable at a time. Each of these sections consists of a serial presentation of studies executed in the area described by the section title. Thus, studies are not presented in the order in which they were undertaken; consequently, it is not unlikely that the reader might be confronted with a comprehensive study at one point and then later the simple study on which the former was based. When appropriate, comments are made which tend to integrate the studies. Broad comments and an overall summary constitute the last section of this report entitled "Conclusions, Discussion, and Recommendations".

Reports of studies are not all equally comprehensive. Some studies have been published or are in press; so the reader is referred to the full accounts which appear or will appear elsewhere. Other studies are dealt with more extensively.

The reader should bear in mind that the investigators were doing nothing more than exploring the possibility of administering tests by projected methods, a touch-and-go activity at best. The focus was on the identification of basic variables. The question, "What values should be assigned the variables so that the test would yield results comparable to those obtained under normal administrative conditions?" guided many studies. Because of this emphasis, and the collateral concern about conditions under which the testing procedure could be applied, the studies reflect both the theoretical and practical.

The writers wish to acknowledge their gratitude and indebtedness to the California Test Bureau, Educational Testing Service, Harcourt, Brace and World, Inc., the Psychological Corporation, and Science Research Associates for their permission to adapt some of their tests for experimental usage. Without their cooperation, the writers would have been indeed handicapped in the prosecution of this project.

II. PROCEDURES AND APPARATUS

This section describes the procedures and apparatus that were common to most of the studies. These matters are discussed here to eliminate repetitious descriptions of them in the context of each study. If an exceptional procedure and/or apparatus was an integral part of a study, then it is described fully during the presentation of that study. However, this section is not a cut-and-dried description of procedures and apparatus.

Comment is directed to why a particular method or piece of equipment was used and, when appropriate, how it seemed to work in practice. Occasionally suggestions are made about what probably would have been more adequate procedures and equipment.

The investigators were confronted initially with the choice of a medium by which to project test materials. The following media were considered at some length—television, motion picture projection, still picture projection of the slide and filmstrip types. Television was discarded reluctantly. The costs, especially those involved in production of materials, of televised presentation were prohibitive. Furthermore, preparing materials for televised presentation would cause them to be cast in permanent form thereby precluding a simple adaptation of them for subsequent testing. For example, if our problem were the investigation of the effect on performance of the arrangement of item difficulties within a set of items, as was actually one of our problems, then we would have been obliged to prepare several tapes each containing the items arranged in a different order. To do so would be expensive. However, we might have purchased portable television broadcasting equipment which would have permitted us to put items on cards and to televise pictures of the cards. This alternative was discarded on the bases of cost of equipment and inconvenience in transporting the transmitter to the testing locale.

To project materials with a motion picture projector was a more feasible alternative. It was discarded principally because the images to be projected were "stills" and it seemed wasteful to devote several hundred frames of 16 mm film to present one inanimate item. But rejecting it was difficult, as was true, too, of rejecting television, for it would permit uniform control of inter- and intra-item exposure intervals from one testing session to the next. Expense was the overriding factor. Since both televised and motion picture presentation permit a new dimension in test materials, animated test items, their rejection was made even more reluctantly.

The choice was reduced to still projection either with a filmstrip or slide projector. Filmstrip was discarded for a number of reasons. Professional photographic reproductions were expensive and through previous experiences we were aware that processing time was unduly long. Thus to choose filmstrip was tantamount to agreeing to photograph and process our own materials. Several trial runs with personal photographic equipment led us to believe that only with expensive and complex equipment could we hope to produce a strip that would consist of frames having uniform quality. It is needless to add that filmstrip, too, is as unadaptable as is the motion picture and television tape. Consequently our choice was slides and a slide projector. Unfortunately the foregoing account makes it appear that we were forced into this decision; the implication is more apparent than real.

Preparing and projecting individual slides had many advantages over other modes which might have been adopted. Perhaps some of the advantages which are cited subsequently are unique to the work situation of the investigators rather than to inherent qualities of the apparatus and methods.

We chose to produce the slides ourselves. We used a Polaroid 800 camera, a Polaroid copymaker, and Polaroid Type 46 Projection Film. This equipment permitted us to produce a slide in about one minute after "assembly line" procedures had been devised and perfected. Slide quality was comparable for our purposes to that obtained with usual 35 mm copy film. Some tendency for slides to fade was noticed but it was not serious enough to necessitate making new slides except in a few cases. Whether fading is a characteristic of Polaroid materials or is attributable to faulty technique is not known, but the investigators lean to the latter view. Second, the necessary slides, projector, and screen were portable and adaptable to nearly all situations. Another advantage, although held in common with modes which were rejected, was the projected screen image could be made to conform with the format of the television screen. A conspicuous advantage was that the slides were easily adapted to varied administrative conditions of testing. When arrangement of item difficulty was studied, it was possible to re-number items easily by use of Para-Tipe, thus permitting the same items to be used in various serial arrangements. Similarly, when the intra-or inter-item exposure interval was altered under different testing conditions, it was accomplished simply. This gain could not be realized by use of any other method.

Slides were projected either with a Polaroid 610 2 1/2 x 2 1/2, 500 watt slide projector or with a Bell and Howell Focustronic 35 mm automatic, 500 watt slide projector. Inasmuch as the Polaroid projector was not automatic, projecting the slides which constituted a test was a two-man operation. One person loaded the slide carriage and put the slide in view at the proper time. The other person emptied the slide carrier and kept time, since intra-and inter-item exposure intervals were rigidly controlled. Coordinating the work of two operators and developing the manual skills were quickly achieved after a short training period.

This method of projecting slides was adequate for early studies but not for later studies which required extremely short exposure intervals and synchronized audio presentation of the test materials. The Bell and Howell projector was used for such studies. Because it accepts 35 mm slides, where the Polaroid accepts 2 1/2 x 2 1/2 slides, it was necessary to cut down the 2 1/2 x 2 1/2 slides which had been used earlier so that they could be put in slide frames for the Bell and Howell. Reducing slides to the new and smaller format required re-making some slides. When slides were made with the Polaroid equipment specifically for the 35 mm projector, the field on the Polaroid Copymaker was adjusted so that the resulting slide would fill the entire 35 mm size frame after being trimmed.

The synchronized audio-visual presentation of materials was accomplished by synchronizing the Bell and Howell projector and the Bell and Howell tape recorder by means of a Kodak Cavalcade Programmer. By means of the programmer, high frequency "blips" were placed on the magnetic tape so that when the tape, which also contained the audio rendition of each test item, was played, the slide changer of the projector was activated, thereby putting a new slide in view. Thus, the item was read from tape when the item appeared on the screen. In each instance, the entire item, stem and alternatives, was read except when the alternatives consisted of rather meaningless materials whose value lay in designating a response which appeared elsewhere, e.g., and item dealing with map reading might inquire about the location of Jonestown on the map which was in view and the item responses consisted of, say, alternative 1, K; alternative 2, L; etc., where K, L, etc., were printed on particular locations of the map.

In all studies except when noted, Ss indicated responses on standard IBM answer sheets. The illumination in each testing room was kept at such a level that response marking, scratch work, etc., were done in comfort. In fact, the illumination level was comparable to that afforded by a dull, cloudy day. Such a description is unscientific but it is informative.

All test directions and preliminary instructions were presented to subjects by projecting them on the viewing screen. Test directions for use with the booklet form of a test sometimes had to be modified, e.g., the typical instruction, "Do easy items first," had to be discarded except in the case when items were presented in triads. Supplementary directions were always presented which described in detail the new mode of testing and the new demands which it entailed. Two points in these directions bear elaboration. First, Ss were informed that each item or each group of items would appear on the screen for only a limited period of time during which they would have to solve and respond to the item. Several practice items preceded each test to familiarize Ss with the testing procedure and with the item pacing which characterized this mode of testing. Generally, practice items were given an intra-exposure interval identical with that which would occur during the actual testing if items were to receive constant exposure; practice items were given the modal intra-exposure interval if intra-item exposure intervals were variable throughout the test. Ss were never informed of item exposure intervals in terms of seconds-of-exposure during the time the test was in progress. Second, Ss were advised in the preliminary directions that each item in the test would be presented one and only one time during the test and, as a consequence, if a response was to be given to an item, then it would have to be given at the time when the item appeared initially.

If a test consisted of separate sub-tests which required separate directions when administered in the usual way, then these directions were presented at the time of presenting those subtests under the projected conditions.

III. INTELLIGENCE

The studies reported in this section deal with adaptations of intelligence tests for projected administration. The study which is presented first was the first investigation which was undertaken in the entire series dealing with the possibility of projected testing. The studies presented in this section are identified, at least in topic heading, by the name of the test which was used to gather the experimental data. Although the published tests were not the object of study, their titles serve as convenient tags for the various studies.

The investigators wish to state that the majority of published tests which served as experimental tests were chosen on the basis of anticipated problems which they posed for projected testing and/or the audiences for which they were intended. Studies reported in this section are based on the School and College Ability Test, the Primary Mental Abilities Test, and the Concept Mastery Test. The form of one of those which was chosen for use contained quantitative items which we felt to pose particular problems for projected testing. Furthermore, it is a widely used test and is part of the battery of the Florida State-Wide Ninth-Grade Testing Program. One of the others is a high-level ability test which can be used with graduate student populations, thereby extending the range of the studies in the sense of age-levels with whom the testing method was used. The other test posed problems because of its timed sub-tests which prevented simple, a priori determination of intra-item exposure intervals as could be done with rather pure power tests. In addition to basing choices on points like those described above, the investigators attempted to choose tests when other factors were equal, which currently enjoy wide usage and which would give major test publishers representation in these studies. Although a paragraph similar to this one does not appear at the beginning of each subsequent section, it might very well appear there, for points such as those described above served as the basis for all test selection.

SCHOOL AND COLLEGE ABILITY TEST[1]

Problem

The general problem was the determination of changes in test scores for a group of Ss who were administered the same test under three sets of administrative conditions. These three sets of conditions are as follows: One administration of the test was done in conformance with normal procedures, i.e., Ss were given test booklets, answer sheets, etc. This administration is referred to as the control administration. The other two sets of administrative conditions were experimental and involved projecting test items on a screen as the method of presenting the test to Ss. One experimental administration consisted of presenting test items one at a time; it is referred to as the "single" presentation. The other experimental administration involved presenting the items three at a time; it is referred to as the "triad" administration. More explicit descriptions of the experimental conditions appear subsequently.

The following comparisons were made of the scores obtained under these three sets of conditions: means, frequency of response, guessing, and the factor patterns of each set of scores.

---

[1] Curtis, H.A. and Kropp, R.P. "A comparison of scores obtained by administering a test normally and visually." Journal Ex. Ed. V. 29. No. 3 (March, 1961), pp. 249-260.

Procedure

Sample. The sample consisted of a ninth-grade class of white students. Complete sets of test data were collected from twenty-nine of the class members. Four students from whom incomplete test data were gathered were discarded prior to the analysis of the data.

Tests. The following tests were administered to *Ss* under normal conditions:

School Ability Test, Form 3A, of the Cooperative School and College Ability Test; Iowa Test of Educational Development, Form X3S, Tests 3-7; Test 3. Correctness and Appropriateness of Expression; Test 4. Ability to do Quantitative Thinking; Test 5. Ability to Interpret Reading Materials in the Social Sciences; Test 6. Ability to Interpret Reading Materials in the Natural Sciences, and Test 7. Ability to Interpret Literary Materials; Thurstone Temperament Schedule; Gordon Personal Profile; Gordon Personal Inventory; SRA Primary Mental Abilities, ages 11-17; Iowa Silent Reading Test, New Edition, Advanced; and Clerical Speed and Accuracy of the Differential Aptitude Test.

The School Ability Test was administered normally and experimentally. All other tests were administered to provide data on the basis of which to evaluate the possible differences between the results obtained from the normal and experimental administrations of the School Ability Test.

The School Ability Test, which was used as the experimental test, consists of four parts; Parts I and III assess verbal abilities, and Parts II and IV assess quantitative abilities. On the basis of item analysis data supplied by the Educational Testing Service, the test was divided into two half-tests, thus permitting the use of a half-test for each set of experimental conditions and thereby eliminating item overlap between the two experimental administrations. Data dealing with the comparability of these half-tests appear in Table 1. The column entitled *N* contains the number of items in each part of the half-test.

Table 1

Comparison of Half-Tests Devised from the School Ability Test

	Form A				Form B		
Part	$\bar{p}$	Sigma	p range	N	$\bar{p}$	Sigma	p range
I	.48	.24	.14-.91	15	.48	.20	.07-.75
II	.47	.24	.09-.91	13	.47	.23	.09-.84
III	.45	.21	.11-.82	15	.45	.23	.14-.95
IV	.44	.23	.06-.80	13	.44	.24	.08-.80

Experimental Conditions. There were similarities and differences between the experimental conditions under which the half-tests were administered on two occasions. The similar conditions were as follows: no inter-item exposure interval, all items were five-choice multiple choice, the total testing times devoted to item solution were the same, directions were the same, and the mode of item presentation was visual.

The chief differences between the experimental administrations were the number of items presented per exposure, the intra-item exposure interval, and the order in which the items were presented.

The single item administration presented one item per exposure. The exposure duration was determined by dividing the number of items included in each part of the test into the amount of time the test authors allotted for that part of the test. Since the number of items and time allowed differed from part to part of the test, the exposure interval was not constant throughout. The following durations were used: Part I, 30 seconds; Part II, 48 seconds; Part III, 20 seconds; and Part IV, 60 seconds. The items within each part of the half-test were presented in the same order in which they appeared in the parent test, thus the item order was that of ascending difficulty.

The triad administration presented three items per exposure. The exposure interval for each part was three times that allowed for the single item administration. The item order within each part was determined in this way. Triads were constructed to have the same mean difficulties and ranges, insofar as possible. Within a triad, items were arranged from least to most difficult. Since it was impossible to make the range of difficulty for each triad exactly the same, the triads were arranged from greatest to least range. In short, the order of items within a triad presented a rhythm of difficulty--simple, average, difficult --which was repeated throughout the entire test.

The reader will note that thirteen items are contained in each of Parts II and IV of the test. Actually fifteen items were used in each so that there would be five triads. Two were borrowed from the comparable part of the other form. However, these borrowed items were not scored.

Testing Schedule. The tests were administered in two sessions, the first preceding the second by a period of three months. The first session included the normal administration of the School Ability Test and the five tests of the Iowa Tests of Educational Development. These tests were administered in half-day sessions on three consecutive days.

The remaining tests were administered over a three-day period as follows: First day, morning--Gordon Personal Profile and Gordon Personal Inventory; afternoon--single item experimental half-test. Second day, morning--SRA Primary Mental Abilities; afternoon--triad experimental half-test. Third day, morning--Thurstone Temperament Schedule, DAT Clerical Speed and Accuracy, and Test I of the Iowa Silent Reading Test.

Experimental Directions. Special directions and sample items were used when administering the experimental test. The directions were distributed, one copy to each S, and were read aloud by the examiner while the Ss read them silently. Prior to the presentation of each of the four parts of each experimental test, special directions were given and a sample item was exposed on the screen. Ss were at no time informed of the amount of time they could spend on each exposure. They were informed that no questions could be asked during the administration of any part of the test.

They were informed in the instructions that they would not be able to refer back to items which they might not have answered when they had been presented.

No unusual circumstances arose during the administration of the experimental tests.

Results

Comparison of Means. The answer sheets from the normal administration of the School Ability Test were re-scored several times to gain scores over items which appeared in the half-tests. In accordance with customary usage of the School Ability Test, Parts I and III were combined to provide a verbal score and Parts II and IV were combined to provide a quantitative score.

The means and standard deviations are presented in Table 2. In this table, and hereafter, the following code is employed: C, control; X, experimental; V, verbal; Q, quantitative; S, single; and T, triad.

Table 2

Means and Standard Deviations on Half-tests from SAT, 3A

	Mean	S.D.
CVS	17.068	5.98
CVT	16.757	6.00
CQS	15.619	3.62
CQT	14.930	3.86
XVS	17.930	6.05
XVT	18.102	6.92
XQS	15.378	3.64
XQT	15.550	3.74

N = 29

Because all data were gathered from the same Ss, it was necessary when computing t tests of mean differences to make corrections for correlated performances. The relevant intercorrelations appear in Table 3 and the t tests appear in Table 4.

The following generalization might be tentatively drawn from the tabular data. The ST administrative conditions lead to higher scores than do the XS conditions. On the verbal section, either experimental condition produces higher mean scores than does the control administration. But apparently no generalization can be drawn for the quantitative items:

Table 3

Correlations between Scores on Control and Experimental Half-Tests of SAT, 3A

	CVS	CVT	XVS	XVT	CQS	CQT	XQS	XQT
CVS		.879	.935	.864				
CVT			.869	.865				
XVS				.854				
XVT								
CQS						.815	.789	.687
CQT							.650	.703
XQS								.753
XQT								

N = 29

Table 4

t Tests of Mean Differences Corrected for Correlated Performances

	CVT	XVS	XVT	CQT	XQS	XQT
CVS	.55	-2.03	-1.59			
CVT		-1.99	-2.07*			
XVS			- .25			
CQS				1.60	.53	.12
CQT					-.76	-.73
XQS						-.25

N = 29

*$p < .05$

CQS and CQT means bracket the counterpart X means despite the fact that the CQS and CQT means should be equal on account of the method used to build the half-tests.

The tendencies described above must be regarded skeptically because the *t* tests revealed no significant differences between the means except in one case. Furthermore, one might attribute the apparent, although not significant, superiority of XT to XS to the fact that XT data were gathered subsequent to the XS data thereby possibly reflecting practice effect.

In summary, the data to this point do not generally indicate, at the .05 level, differences between mean scores which are functions of administrative conditions. However, the control data were collected three months prior to the collection of experimental data. Nevertheless, the intercorrelations which appear in Table 3 indicate that *Ss* would fall in roughly the same order on the experimental test as they were ordered by the control test.

Comparisons by Item Types. The preceding analysis dealt with means on the Q and V sections of the test inasmuch as Q and V scores are generally reported when the School Ability Test is used. However, the Q and V scores, respectively, are established by summing success over two parts of the test each of which has relatively distinct kinds of items. Sample items illustrate this difference.

Verbal

Part I We had worked hard all day so that by evening we were quite (   ).
A. small B. tired C. old
D. untrained E. intelligent

Part III. Chilly

A. tired B. nice C. dry
D. cold E. sunny

Thus, the items in Part I contain a sentence as a stem, whereas in Part III the item stem is merely a stimulus word.

Quantitative

Part II 5413 A. 586 B. 596 C. 696
-4827 D. 1586 E. None of these

Part IV. Four $10-bills are equal to how many $5-bills?
A. 20 B. 10 C. 8
D. 40 E. 2

Thus, the items in Part II consist of stems containing a set of numbers and an operations sign, whereas the items in Part IV consist of stems which are statements of word problems involving arithmetic.

Analyses were done to determine whether differences existed between success on particular item types in association with different conditions under which the items were administered. The relevant data appear in Tables 5 (means and standard deviations), 6 (intercorrelations) and 7 (*t*'s).

Table 5

Means and Standard Deviations of Part Scores on Half-Tests from School Ability Test

Condition	Verbal					Quantitative			
	S*(I)		NS**(III)			S(IV)		NS(II)	
	X	sigma	X	sigma		X	sigma	X	sigma
CVS	8.52	3.89	8.55	2.94	CQS	6.48	2.01	9.14	2.11
XVS	9.24	2.98	8.69	3.28	XQS	7.24	1.79	8.10	2.26
CVT	8.45	2.96	8.31	3.46	CQT	6.45	2.01	8.48	1.45
XVT	8.79	3.97	9.31	3.51	XQT	8.24	2.18	7.31	2.37

N = 29

* stem

** no stem

Table 6

Correlations Between Part Scores on Half-Tests From School Ability Tests

Verbal			Quantitative		
Variable Pair	S(I) r	NS(III) r	Variable Pair	S(IV) r	NS(II) r
CVS-XVS	.77	.86	CQS-XQS	.58	.71
CVT-XVT	.76	.79	CQT-XQT	.59	.48
XVS-XVT	.71	.76	XQS-XQT	.54	.41

N = 29

Table 7

t Tests of Differences Between Correlated Means of Part Scores from Half-Tests of School Ability Test

Verbal			Quantitative		
Variable Pair	S(I) *t*	NS(III) *t*	Variable Pair	S(IV) *t*	NS(II) *t*
CVS-XVS	-1.55	-.43	CQS-XQS	-2.29*	3.28*
CVT-XVT	- .71	-2.36*	CQT-XQT	-5.03*	3.00
XVS-SVT	.846	-1.40	XQS-XQT	-2.73*	1.67

N = 29
*$p < .05$

The tabular data reveal for the items in the two verbal parts of the test that regardless of whether the item contained a verbal statement or not, the experimental administrative procedures led to slightly higher mean scores than did the normal administrative conditions. Whether this is due to the effect of the experimental conditions or the fact that the experimental data were collected three months after the control data is very much in doubt. On the quantitative items a differential in response accuracy is associated with method of test administration. On the quantitative items having statements for stems, the experimental administrations led to higher mean scores than did the control administration. However, on the quantitative items which had for a stem a series of numbers and an operation sign, the control means were significantly different from and greater than the experimental means. This finding is worthy of comment.

It is felt that the reduced mean scores on the quantitative no-stem items is attributable to any or all of the following, since the problems were so involved the examinee probably copied the data from the screen to his scratch paper in order to solve it, thus (1) reducing his working time due to the time involved in copying, (2) committing clerical errors when copying, (3) after solving the problem, needing to locate a similar response on the screen during which errors in remembering might have occurred. However, it should be noted that non-response to these items was significantly less than the amount of non-response on the same items when administered normally. Whether the poorer performance is due simply to guessing or to factors noted above is problematical.

Another factor worthy of specific mention is that the intercorrelations of the quantitative items with regard to method of administration are of a lesser magnitude than comparable intercorrelations for the verbal items. Generally the lowest intercorrelations are based on the no-statement quantitative items.

Analysis of Omits. It was hypothesized that *Ss* would be more likely to give a greater number of responses to the items when experimentally presented than when normally presented simply because all items would be exposed to them on the experimental administration whereas on the normal administration their test taking habits might have caused them to omit some items. Also it was hypothesized that there would be no difference in response frequency between the triad and single experimental presentations. Both of these hypotheses held at the .05 level of confidence. The Wilcoxon Matched-Pairs Signed-Ranks Test was used in computing the significance of differences. The pertinent data are presented in Table 8.

Table 8

Comparison of Non-Response on Experimental and Control Administrations

Non-Response Hypothesis	Significance Level
XVT = XVS	No difference
XVT < CVT	.025
XVS < CVS	.01
XQT = XQS	No difference
XQT < CQT	.005
XQS < CSQ	.005

N = 29

It appears that the experimental test administration procedure leads to a greater number of item responses than does the normal administrative procedure during which the examinees are responsible for their own pacing.

Although differences in non-response are associated with mode of test administration, the data were analyzed to determine if non-response was related to certain personality variables on one kind of administration but not on the other. A response accuracy score was calculated for each S̲ on the basis of his responses on the XQS and CQS administrations, separately. The accuracy score for S̲ was equal to the ratio of his "rights" score to the total number of items he responded to. The score was computed in this way to discount differences in ability which would affect the S̲'s probability of non-response. The personality variables which were evaluated with regard to response accuracy were "Cautiousness" from the Gordon Personal Inventory and "Reflective" and "Impulsive" from the Thurstone Temperament Schedule. A 2 x 2 Chi square analysis produced the results shown below.

CQS response accuracy

vs Cautious - $\chi^2 = .00$ $p > .05$
vs Reflective - $\chi^2 = 2.28$ $p > .05$
vs Impulsive - $\chi^2 = .56$ $p > .05$

XQS response accuracy

vs Cautious	- $\chi^2 = .65$	$p > .05$
vs Reflective	- $\chi^2 = .00$	$p > .05$
vs Impulsive	- $\chi^2 = .00$	$p > .05$

The data indicate that tendency to omit items is not related to these three personality variables on either the normal or experimental administrations.

Intercorrelations. The intercorrelations between control and experimental administrations of the experimental tests appear in Table 9. The pertinent correlations are those which are based on scores over the same items which were administered under experimental and control conditions. For the verbal part of the test, CVS and XVS correlate at .93, and CVT and XVT correlate at .86; thus, indicating rather high association between scores collected under control and experimental conditions of test administration. If these intercorrelations are regarded as equivalence coefficients, then it is proper to adjust them for a test twice as long since they are based on half-tests. Correction by the Spearman-Brown formula produce the following correlations: CVS and XVS, .96; and CVT and XVT, .92. For the quantitative part of the test, CQS and XQS correlate at .79, and CQT and XQT correlate at .70. When these coefficients are raised for a test twice as long, then the coefficients become .88 and .82, respectively. The adjusted correlations for the quantitative section are relatively high but probably indicate an unsatisfactory level of reliability. The difference in magnitude between the correlations based on the verbal part of the test and the correlations based on the quantitative part of the test is probably due in part to the difficulty described earlier which the Ss encountered when attempting to solve quantitative items when presented visually.

Table 9
Intercorrelations of Control and Experimental
Administrations of Experimental Test

	CVS	CVT	XVS	XVT	CQS	CQT	XQS
CVT	.88						
XVS	.93	.87					
XVT	.86	.86	.85				
CQS	.74	.71	.66	.59			
CQT	.62	.66	.61	.53	.81		
XQS	.65	.57	.55	.55	.79	.65	
XQT	.57	.55	.57	.57	.69	.70	.75

N = 29

Factor Matrix. All Ss were administered the experimental test under normal and experimental conditions. In addition, they were administered Tests 3-7 of the Iowa Tests of Educational Development, the Thurstone Temperament Schedule, the Gordon Personal Profile, the Gordon Personal Inventory, the SRA Primary Mental Abilities, the Iowa Silent Reading Tests, and the DAT Clerical Speed and Accuracy Test. These were administered to provide data on the basis of which to determine if the experimental administrative procedures tended to produce differences between the factor content scores from the experimental test when administered under control and experimental conditions. The correlation matrix was factored by the centroid method and the factor matrix after a quartimax rotation appears in Table 10. No noteworthy differences were noted between the factorial content of the experimental test administered under normal and experimental conditions.

Table 10

Rotated Factor Matrix*

Tests	I	II	III	IV	V	VI	VII	VIII
01 CVS	9470	-0593	-0261	0279	-0622	0450	1028	-0497
02 CQS	7705	0607	-0826	-0001	-0741	-5216	1367	-1388
03 CVT	9382	0903	-0619	-0032	-1244	0695	0778	0302
04 CQT	6866	-0693	-1172	0164	-0942	-5056	-0170	2789
05 Control Total Score	9616	-0272	-0680	-0206	-0950	-1538	0797	0447
06 XVS	9172	0117	0096	1013	-0833	1002	2484	0740
07 XQS	6916	-2285	1308	-1153	0030	-5205	0192	-1300
08 XVT	8979	0640	1574	-0131	-0206	1715	0254	0266
09 XQT	6983	-0933	0288	-2227	0233	-3572	-0874	1214
10 Gordon Ascendancy	1945	-8023	-1591	0575	-0143	-0468	0624	-0155
11 Gordon Responsibility	2244	-0709	7513	-1007	-1000	0539	2756	-0272
12 Gordon Emotionally Stable	-0769	2919	5357	-0118	0053	1544	3273	0488
13 Gordon Sociability	-0349	-8326	0611	0615	0541	-2651	0428	-1051
14 Gordon Cautiousness	0177	-1044	8164	-0262	1668	0581	-1055	0201
15 Gordon Original Thinking	3424	-3108	6156	0625	-1326	-0842	1408	0299

(Continued on next page)

Table 10 (continued)
Rotated Factor Matrix

	Tests	I	II	III	IV	V	VI	VII	VIII
16	Gordon Personal Relation	0506	1643	7077	0675	0727	-1656	-3099	-0778
17	Gordon Vigor	3675	-1835	2388	-1200	-1771	0848	1920	-3291
18	Thurstone Active	2162	-5710	-2405	1607	-3380	-0101	0045	-1353
19	Thurstone Vigorous	0141	-2368	0327	6066	0033	2479	-0151	0150
20	Thurstone Impulsive	-0334	-5266	1619	-1468	-0926	0039	4332	-1165
21	Thurstone Dominant	0551	-8274	-0373	-2266	0864	1955	-1445	1761
22	Thurstone Stable	0084	-3282	2241	-6277	-0572	1103	-0068	-0376
23	Thurstone Sociable	-2000	-8546	2144	0256	-0628	0359	-0068	0164
24	Thurstone Reflective	1779	-0469	0887	3048	1842	1023	1308	4205
25	PMA Verbal Meaning	8326	1297	0700	2128	2912	1519	0246	-0742
26	PMA Space	3468	-0397	-4441	2464	1892	1527	2155	1830
27	PMA Reasoning	4617	0644	-1080	-3836	4377	0240	-0485	1625
28	PMA Number	6178	-2343	-3273	-0321	1393	1792	1419	-3654
29	PMA Word Fluency	6544	-0806	2111	0169	0016	-1743	1670	-3978
30	PMA Scholastic Aptitude	8401	0214	0320	1102	4630	1406	0120	-0016
31	DAT Clerical Sp. and Acc.	3204	-0672	0410	1342	6190	0276	-0068	-0371
32	Iowa Silent Reading Rate	5785	-0459	0503	3930	1196	0843	-2257	1481
33	Iowa Reading Comprehension	8074	1520	0140	0178	1090	-2649	0221	1216
34	ITED Correct Expression	8864	1020	1034	-2423	0852	1187	0409	-0125
35	ITED Quantitative	7562	-1799	-0538	0121	-0746	-3008	-1130	-1024
36	ITED Social Studies	8527	0642	0200	-0307	-0138	0829	-4265	0980
37	ITED Natural Science	8050	0542	1314	-0053	-0844	0956	-3647	-1054
38	ITED Literary Material	8681	0084	-0433	-0720	-0921	1863	-3272	0304

*All decimal points omitted.

Observations During Experimental Testing. Ss reactions to the experimental testing and E observations are briefly summarized below. Ss reported no special difficulty in dealing with the experimental administration of items from the verbal parts of the School Ability Test. However, they reported difficulty in solving items from the Quantitative sections. They reported, and E observed, that after solving a quantitative item, calculating on scratch paper, they would view the screen to locate the correct alternative only to find that a new item had taken the place of the item they had been solving. Ss reported this to occur frequently and they claimed the experience was frustrating. These reactions were particularly noticeable on the single-item presentation, but less so on the triad presentation. Of the two experimental conditions, Ss indicated a preference for the triad presentation.

Ss appeared to be restless and inattentive when presented easy items. This is probably due to the fixed amount of time allotted to each item regardless of difficulty. Since no audible signal was given to indicate the removal of an old item and the presentation of a new item on the screen, Ss who were not paying attention to the screen sometimes noticed the new item only after it had been on the screen for several seconds.

There were other reactions, but those above are the noteworthy ones.

Summary

In the main, the attempt to reproduce by projected presentation the test results obtained by normal presentation was successful. Certain difficulties encountered by Ss in working from screen to paper to screen were identified. There did not appear to be important differences in performance between normal and experimental conditions which were associated with certain personality attributes. The factorial content remained substantially constant.

Both the analysis of Ss performance and the E's observations suggest that television testing would permit a greater test coverage of an area within a given unit of time than is commonly attained by traditional methods. This is true both because variable exposure intervals which are functions of item difficulty and type will probably lead to greater efficiency and because the method appears to carry Ss further through the total test.

TERMAN CONCEPT MASTERY

Problem

The purpose of this study was to determine the effect on test scores which might be attributed to differences in intra-item exposure intervals, inter-item exposure intervals, and their interactions.

Procedure[1]

Subjects. Ss consisted of all members of eight classes of graduate students; however, students for whom English was a second-language were eliminated prior to analyses of data. No S was a member of more than one classroom group. No data were collected about the comparability of these groups prior to experimentation.

Experimental test. The Terman Concept Mastery Test, Form T, was adapted for use in this study. Part I of the test consists of 115 word pairs for each of which the S indicates whether the pairing is of synonyms or antonyms. Part II consists of 75 analogies for each of which S selects one of three alternatives to complete the analogy. The test was sub-divided into three experimental forms by assigning items to forms in this manner--Item 1 to Form A, 2 to Form B, 3 to Form C, 4 to Form A, etc.

Prior to this assignment item 115 of Part I was omitted so that items per form would be equal. Each form, i.e., third-test, was prepared in booklet form comparable in all details to the booklet form of the parent test. These forms were designated T (Terman) A (form of third-test) C (for control administration), TBC and TAC, respectively. Each of these was duplicated on slides, one item per slide. These forms were designated TAX, TBX, TCX, respectively. Thus, TAC and TAX consisted of the same items. The equivalence of the control administration forms was determined by administering them in random order to twenty-seven graduate students. Pairs of means, after correcting for correlated performance, did not differ at the .05 level; variances did not differ at the .05 level; intercorrelations ranged from .784 to .848; and there were no differences in mean test-taking times (control administration was untimed) between forms on each of Part I, Part II, and the total test, respectively. It was then assumed that the experimental forms were equivalent.

Method. Ss in each group were administered the three experimental forms of the test--two administrations were under projected conditions and the other was under normal conditions. In each group, the control administration was administered last. Directions, sample items, etc., were projected during the projected administrations. Ss were not informed about the length of exposure interval on projected administrations; they were told that each item would be presented only one time.

The major independent variables that were studied were the intra-item exposure interval and the inter-item interval. For some administrations, the value given the latter variable was zero. Since Part I and Part II of each form required different working times as revealed by preliminary control administrations and since each Part contained a different item type, different exposure intervals were set for each Part. An experimental condition is indicated, e.g., by 12-0, 24-0, which indicates each item of Part I was exposed for 12 seconds, each item in Part II was exposed for 24 seconds, and the duration of the inter-item interval for each Part was zero. A schedule of conditions appears below.

---

[1] A preliminary report of this study will appear in AV Communication Review. Curtis, H.A. and Kropp, R.P. "A comparison of standard and visual administrations of the Concept Mastery Test." AV Communication Review. In press.

Group	Form	Condition	N
I	TAX	12-0, 24-0	16
	TBX	6-0, 12-0	
	TCC	Normal	
II	TBX	8-0, 16-0	19
	TCX	4-0, 8-0	
	TAC	Normal	
III	TBX	8-4, 16-8	12
	TAX	12-0, 24-0	
	TCC	Normal	
IV	TBX	6-4, 12-8	13
	TCX	10-0, 20-0	
	TAC	Normal	
V	TAX	3-0, 6-0	20
	TCX	5-0, 10-0	
	TBC	Normal	
VI	TAX	10-4, 20-8	15
	TCX	4-4, 8-8	
	TBC	Normal	
VII	TAX	3-4, 6-8	14
	TCX	5-4, 10-8	
	TBC	Normal	
VIII	TAX	12-4, 24-8	10
	TBX	8-4, 16-8	
	TCC	Normal	

Analysis. Nine scores were obtained for each S from each form of the test. Scores consisted of Part I, Part II, and total scores for the number of right answers, the number of wrong answers, and the number of omitted items.

The data were analyzed through the use of a multiple regression technique. Binary variables were generated for each level of each independent variable. Thus eight variables were generated for the eight groups of *Ss*. If, for example, *S* were a member of Group III, he received a score of one for the Group III variable and zeroes for all other Group variables. Binary variables were also generated for three orders, two tests, eight exposure conditions, one inter-item exposure time and seven condition by inter-item exposure time interaction variables. Thus the total number of independent variables was twenty-nine. Correlation matrices were computed for the twenty-nine independent and nine dependent variables. Multiple correlations were calculated using the integrative technique developed by Ward and Bottenburg.[1] The nine total R^2's (twenty-nine variables used as predictors) were subjected to tests of significance. *F*-ratios for the various main effects and the interaction were computed through the use of the formula which appears below.

$$\underline{F} = \frac{R_b^2 - R_s^2}{1 - R_b^2} \times \frac{(N-r-1)}{t} \quad \text{with df} = t \text{ and } (N-r-1)$$

In the formula R_b^2 is the square of the multiple correlation of the dependent variable with all the independent variables, R_s^2 is the square of the multiple correlation of the dependent variable with the subset of independent variables that excluded the effect under consideration, N was the total number of observations (358), r was the dimension of the space generated by the total set of independent variables and t was the reduction of the dimension of the total space resulting from subtracting the dimension of the subset of independent variables of the effect under consideration.

Results

Table 11 contains the *F*-ratios, computed by the technique described above, for the three dependent variables whose total R_s^2 were significant. These variables were Part I rights, Part II rights, and Total rights. It can be determined from the tabular data that none of the main effects or interactions for any of the dependent variables produced significant *F*-ratios. The failure of even the shortest exposure times to produce scores which differ from the control scores was unexpected. It is possible that the design and analysis of the study was inadequate to detect real differences in the treatment groups. However, it may be that under suitable conditions of motivation and readiness to react, *Ss* can respond with far greater rapidity than is usual for them without sacrificing quality of performance. Further research was undertaken to extend the present findings. The follow-up study is presented subsequently.

---

[1] Statistical Methodology and Analysis Branch. *An Iterative Multiple Regression Procedure Using the IBM 650*. Technical Memorandum, WCLL-TM-60-3, 15 January, 1960, Personnel Laboratory, Wright Air Development Division, Air Research and Development Command, United States Air Force, Lackland Air Force Base, Texas.

Table 11

F-Ratios for Variable Having Significant Total R_s^2

Source	df	Part I Rights	Part II Rights	Total Rights
Total	26	1.56*	1.59*	1.64*
Groups	7	1.49	1.50	1.85
Tests	2	1.12	.15	.13
Orders	2	.00	.00	.09
Conditions	7	.48	1.32	.62
Inter-item Exposure	1	1.67	3.20	1.42
C x I	7	.24	.17	.19

*$p < .05$

TERMAN CONCEPT MASTERY: THRESHOLD TIMES

Problem

This study is an extension of the study presented above which reports no significant differences between mean scores on a high level verbal ability test when administered under normal and various projected conditions. The purpose of this study was to establish a "threshold exposure time," i.e., minimum intra-item exposure interval, for items in such a test.

Procedure

Subjects. Ss consisted of sixty-seven graduate students who were enrolled in six different classes ranging in size from seven to twenty students.

Test. The three published forms of the Concept Mastery Test were subdivided into six equivalent forms. Only the items in Part I (synonyms and antonyms) of the published forms were used in this study. Preliminary studies indicated that the six equivalent forms were equivalent in fact.

Testing Conditions. All forms were presented by projected methods; one item per exposure, item order in accordance with order in the published forms, and no inter-exposure interval. The independent variable was intra-item exposure interval and it was assigned the following values:

Condition	Item Exposure (Seconds)
A	4
B	3
C	2
D	1
E	.5
F	.2

Special directions were prepared and they were read to *Ss* prior to testing. Projected sample items were presented prior to each condition to acquaint *Ss* with the speededness with which items would be presented. Item responses were marked on separate standard IBM answer sheets.

All *Ss* were administered all six forms of the adapted Concept Mastery Tests. Each form was administered under a different condition. Order of administering could conceivable effect scores; consequently order of forms and order of conditions were randomized over the six groups of *Ss*.

Analysis. The results were analyzed using a treatment-by-subject design. The model for the analysis was the following:

$$Y_{ij} = M + T_i + S_j + (TS_{ij})$$

Y_{ij} = Number right on one equivalent form of adapted CMT

T_i = Treatment, intra-item exposure, $i = 1, 2, \ldots 6$

S_j = Subject, $j = 1, 2, \ldots 67$

Results

An analysis of variance was completed using number right on each of the six forms of the adapted CMT as the measured variable. Results of this analysis appear in Table 12.

Table 12

Analysis of Variance

Source of Variation	df	MSS	*F*
Treatments	5	1275.364	223.86**
Subjects	66	34.843	6.12**
(TxS) (Error)	330	5.697	

$^{**}p < .01$

The Newman-Keuls Sequential Range Test was used to analyze the differences between the obtained means for the six treatments. The results appear in Table 13.

Table 13

Between Treatment Differences According to Sequential Range Test[1]

(4-3)	(4-2)	$(4-1)^{**}$	$(4-.5)^{**}$	$(4-.2)^{**}$
(3-2)	$(3-1)^{**}$	$(3-.5)^{**}$	$(3-.2)^{**}$	
$(2-1)^{**}$	$(2-.5)^{**}$	$(2-.2)^{**}$		
$(1-.5)^{**}$	$(1-.2)^{**}$			
$(.5-.2)^{**}$				

---

[1] Numbers in parentheses are pairs of intra-item exposure intervals.

$^{**}p < .01$

Results of the analysis of variance indicate a significant difference between treatments and between subjects. The significant F-ratio for Ss indicates the reliability of the measurement over the several treatments. Substitution of the mean sums of squares in the appropriate formulas for estimating reliability of measurement led to an estimate of .85.

The significant F-ratio for treatments indicates a significant difference in mean performance which is related to the duration of the intra-item exposure interval. The subsequent test of differences between means indicated significant differences between pairs of means except those based on the three longest intra-item exposure intervals. The data indicate that the "threshold" of intra-item exposure interval for items of the type used probably falls somewhere between one and two seconds. The actual existence of such a threshold would tend to substantiate the results of the study previously reported.

PRIMARY MENTAL ABILITIES TEST

Problem

When a test is administered one item at a time by projected methods, a problem arises as regards the method which might be used to establish the intra-item exposure interval. The duration of the interval might be constant over all items, or it might be variable over all items as a function of the difficulty level of each item. The problem is especially serious when the projected method is to be applied to a power-speed test, i.e., a test which is primarily a measure of power but which is sufficiently speeded so that all examinees might not have time to attempt all items. During a projected test all items are presented to examinees; thus simply apportioning total time for normal administration to all items is probably not a feasible method by which to establish exposure intervals for projected testing. The investigation reported here deals with the comparative effects on a power-speed test of assigning intra-item exposure time a constant value and assigning the interval different values according to the several item difficulties.

Procedure

Subjects. Fifty-six white tenth-grade students served as *Ss*. On the basis of the median of their scores on the School and College Ability Test (SCAT), *Ss* were divided into a high and a low group. *Ss* in the high and low groups were then randomly assigned to four treatment groups, each group consisting of seven high scoring and seven low scoring *Ss* on the SCAT. Analysis of variance revealed no between groups differences on the SCAT.

Test. The Primary Mental Abilities Tests (PMA), ages 11-17, Form AM, was used as the experimental test. It consists of four separately timed subtests--verbal, space, reasoning, and number. It might be regarded as a power-speed test.

Testing conditions. The four experimental conditions were as follows: Treatment A: One group was administered the PMA in the usual paper and pencil method in accordance with directions contained in the PMA Test Manual. Treatment B: Another group was administered the PMA in the usual paper and pencil manner except that time limits on the subtests were extended on the basis of a preliminary study which is described subsequently. Treatment C: Another group was administered the PMA under projected conditions, one item per exposure, and the intra-item exposure intervals within a subtest were variable and were based on data obtained from a preliminary study. Treatment D: the fourth group was administered the PMA under projected conditions, one item at a time, and the intra-item exposure interval within each subtest was assigned a constant value on the basis of data obtained from a preliminary study. For all treatments the order of subtests and items within subtests was invariant. For treatments C and D, the projected administrations, there was no inter-item exposure interval; and directions, sample items, etc., were presented orally and visually by means of an electronically coupled tape recorder and slide projector.

The preliminary study mentioned above was based on fifty-two white tenth-grade students who were administered the PMA under normal conditions except that they were permitted to work on each subtest until they finished. Their working times on each subtest were recorded. Mean working times were then computed for each subtest. The answer sheets of these students were then analyzed to determine item difficulty values. The mean time for each subtest, as determined on this extended testing time basis, was used as the time for that subtest under Treatment B. For each subtest separately, items were assigned to three categories--most difficult, average difficulty, and least difficult--on the basis of the item difficulty data. The mean time per subtest as determined on the extended test time basis was set equal to $X + 1.5X + 2X$. Solving for X yields the total time to be allotted to the least difficult items, etc. The variable exposure intervals for Treatment C were set in this way. For Treatment D, the constant exposure interval for each item in a given subtest was determined by dividing the number of items in the subtest into the mean testing time for that subtest as determined on the extended testing time basis. The intervals for each subtest, items, etc., are presented in Table 14.

Table 14

Subtest and Item Times

Subtest	A Normal Time (minutes)	B Extended Time (minutes)	C Variable Item Times (seconds) MD	 AD	 LD	D Constant (seconds)
Verbal	4	7	11	9	6	8
Space	5	9	37	27	18	27
Reasoning	6	10	28	21	14	20
Number	6	16	19	15	10	14

Analyses. The obtained data for each subtest of the PMA were analyzed in a 2 x 4 analysis of variance design which appears below.

Treatments

	A	B	C	D
HS	7	7	7	7
LS	7	7	7	7

Where:

A, B, C, and D correspond to treatments described above.
HS and LS refer to high and low scoring Ss on SCAT.

When the treatment effect was significant, a sequential range test of the means was executed to determine differences between pairs of treatment means.

Results

Results of the analysis of variances appear in Table 15 (Verbal), Table 16 (Space), Table 17 (Reasoning), and Table 18 (Number).

Treatment differences occurred on all subtests except Verbal. Newman-Keuls Sequential Range Tests were computed to determine the significance of difference between pairs of Treatment means. They appear below for each subtest. One asterisk and two asterisks indicate significance at the .05 and .01 levels, respectively.

Table 15

Analysis of Variance of Verbal Subtest Scores

Source	df	MS	*F*	P
Treatment	3	144.87	2.55	NS
SAT	1	611.16	10.77	< .01
Tr x S	3	49.30	.87	NS
Error_w	48	56.76		

SAT Level Means

High	Low
31.07	24.16

Table 16

Analysis of Variance of Space Subtest Scores

Source	df	MS	*F*	P
Treatment	3	992.40	49.79	< .01
SAT	1	21.87	1.10	NS
Tr x S	3	51.07	2.56	NS
Error_w	48	19.93		

Treatment Means

A	B	C	D
27.57	47.93	40.50	37.93

Table 17

Analysis of Variance of Reasoning Subtest Scores

Source	df	MS	*F*	P
Treatment	3	106.31	3.33	< .05
SAT Level	1	77.79	2.44	NS
Tr x S	3	4.88	.15	NS
Error_w	48	31.90		

Treatment Means

A	B	C	D
11.93	18.29	15.71	17.07

Table 18

Analysis of Variance of Number Subtest Scores

Source	df	MS	*F*	P
Treatment	3	4175.57	47.98	< .01
SAT Level	1	274.57	3.15	NS
Tr x S	3	68.43	.79	NS
Error_w	48	87.02		

Treatment Means

A	B	C	D
17.21	47.00	52.57	54.07

Space Subtest

(B-A)** (B-D)** (B-C)**

(C-A)** (C-D)**

(D-A)**

Thus, every Treatment mean differed at the .01 level from every other Treatment mean on the Space subtest.

Reasoning

(B-A)* (B-C) (B-D)

(D-A) (D-C)

(C-A)

Thus, only the normal administration and the extended time administration differ at the .01 level from each other.

Number

(D-A)** (D-B) (D-C)

(C-A)** (C-B)

(B-A)**

Thus, the normal administration differs from the extended time and projected administration but no differences occur between pairs of the latter means.

Over all it would appear that, except on the Space subtest, the extended time limit and the projected administration do not differ with respect to obtained mean scores. Generally, the data show that the mean scores obtained under the normal administrative condition and the extended time limit administration do differ significantly. Generally, variable and constant intra-item exposure intervals on the projected administrations do not seem to be associated with differences in mean scores.

In summary, it would appear that a power-speed test of the kind used in this investigation might be adapted to projected administration by extending the time limits on the normal administration and then equating the results of projected administration to the extended time limit results. However, the investigation does not provide evidence bearing on the question of whether the extended time limit and/or projected administrations lead to factor structures different from the factor structure of the test when administered in the normal way. At any rate, it is obvious that new norms would have to be developed on scores yielded under projected administrative conditions.

IV. ACHIEVEMENT

The application of the projected method of test administration to achievement tests was explored in several studies which are described in this section. The studies do not constitute a sequential investigation of any particular point in projected testing. Only one of these studies does not deal with a standardized, published test. It is discussed first.

GOVERNMENT[1]

Purpose

The purpose was to investigate the effects of two variables--constant and variable intra-exposure intervals, and visual and audio-visual presentation of test items--and their interactions.

Procedure

Subjects. The Ss were sixty-four college students who were registered in two sections of American Government which were being taught by the same instructor. Each section had been taught approximately the same content until the time of the experimental testing. All Ss in each section, separately, were sorted into control and experimental groups, roughly in the proportion of one to two, on the basis of American Council on Education Psychological Examination scores. The Ss for whom ACE scores were unavailable were sorted randomly into the two groups.

Test. The test was a fifty-item multiple-choice test on the content of American Government which is taught at the lower undergraduate level. It was constructed by the course instructor who randomly selected the items from a pool of items having content validity and which had proved to be discriminative on previous administrations. The fifty items were then sorted into three categories on the basis of the number of words in the stem and alternatives. Thereby, reading load was construed to be reading amount. The categorized items were then randomly sorted into two additional categories. All items in one category were assigned item numbers 1-25 on a random basis; the items in the second category were assigned item numbers 26-50. Therefore, the test consisted of two half-tests which were similar as regards reading load. Because item difficulty data were not known in advance, it was not known initially if the half-tests were of comparable difficulty.

Testing conditions. The C group of each section was administered the test under normal conditions by the course instructor.

One X group was administered the test under the following conditions: each of items 1-25 was given a constant exposure, 42 seconds; items 26-50 were given variable exposures--32, 42, and 52 seconds; and each item was read to Ss when it appeared on the screen. The other X group was administered the test under the conditions described above except that items 1-25 received variable exposure intervals, items 26-50 received constant exposures, and there was no audio supplementation of the items.

---

[1] Kropp, R. P., King, F. J., Burr, W. L., and Beard, J. G. "Differences in achievement test scores associated with item exposure time and audio reinforcement." Journal Ex. Ed., in press.

The exposure intervals described above were calculated as follows: *Es*' previous experience that projected testing takes less test-taking time than does normal administration led them to set the maximum projected test-taking time at thirty-five minutes. The 42-second exposure time was equivalent to giving an item its proportionate share of the total time. Variable exposures were set by giving 42-second exposures to each item in the middle category of reading load, 32-second exposures to items in the lowest category, and 52-second exposures to items in the highest category.

There was no inter-item exposure interval. Items were presented one at a time during the projected testing. Within the limits described above, item order was random. *Ss* in the X groups were given test directions, sample items, etc., by projection. The directions stated the length of the constant exposure interval, the lengths of the variable exposure items, and which block of items would have constant exposures and which would have variable intervals. When an item appeared on the screen, *Ss* were not informed of the length of the exposure interval which would ensue. *Es* administered the experimental tests to both groups.

Analysis of data. The following analyses of data were executed: (a) analysis of variance on C scores on items 1-25, and 26-50, separately, to determine differences between half-tests and C groups; (b) analysis of variance of scores from C group (C group from each section combined) and from both X groups on each half-test, separately, and for the whole test; and (c) analysis of variance scores from X groups according to testing condition, item exposure interval, and level of mental ability of *Ss*.

Results

Analysis of variance between C groups and half-tests. The results of the analysis of variance between C groups and half-tests appear in Table 19.

Table 19

Analysis of Control Groups by Half-Tests

Source of Variation	*df*	*ss*	*ms*	*F*
Between Groups	1	96.95	96.95	3.64
Error $_B$	23	612.53	26.63	
Between Halves	1	40.50		12.54**
G x H	1	18.75		5.80*
Error $_W$	23	74.25	3.23	

*$p < .05$
**$p < .01$

They indicate that significant differences occurred between half-tests. The interaction between groups and half-tests was significant. The means appear in Table 20.

Table 20

Mean Scores of Control Groups on Half-Tests

	Half-Tests	
Control Groups	1	2
A (N = 14)	12.86	13.57
B (N = 11)	8.82	12.00

The data in Table 20 indicate control group A was superior, but not significantly so, to control group B, and that items 26-50 were less difficult than items 1-25. Perhaps the significant interaction is due to differences in the instruction of the two groups.

Analysis of variance of C and X groups on each half-test and the whole test. For comparison purposes the two C groups were combined into a single group. Hereafter, C group refers to this combination. The analysis of variance data for the three groups on each half-test, separately, and the whole test appear in Table 21.

Table 21

Analysis of Variance of Groups on Half-Tests and Whole Test

Test	Source of Variation	df	ss	ms	*F*
First-Half	Groups	2	13.91	6.96	.53
	Error	61	808.53	13.25	
Second-Half	Groups	2	18.29	9.15	.62
	Error	61	901.57	14.78	
Whole Test	Groups	2	63.51	31.76	.70
	Error	61	2,782.85	45.62	

No significant between group differences were noted for either half-test or for the whole test. The group means by half-tests and whole test appear in Table 22.

Analysis of variance of scores from X groups by testing condition, item exposure time, and level of mental ability of Ss. Although previously cited analyses failed to reveal between group differences, inspection of the score distributions revealed substantial differences between groups. Consequently,

Table 22

Group Means on Each Half-Test and Whole Test

Test	Groups		
	C (N=25)	X_A (N=15)	X_B (N=24)
First-Half	11.08	12.13	11.96
Second-Half	12.88	13.93	14.00
Total Test	23.96	26.06	25.96

the analysis of variance in experimental conditions was done by ACE level to rid group variation of that variation due to S differences in mental ability. Each X group consisted of five high-scorers on ACE and ten low-scorers. Each group consisted of matched Ss. The results of the analysis appear in Table 23.

Table 23

Analysis of Variance of Scores from Experimental Groups

Source of Variation	df	ss	ms	F
Presentation (P)	1	26.66	26.66	1.87
ACE (A)	1	66.01	66.01	4.64*
P x A	1	4.41	4.41	
Error $_B$	26	369.85	14.22	
Item Exposure (I)	1	.26	.26	
P x I	1	41.67	41.67	5.81*
A x I	1	5.21	5.21	
P x A x I	1	1.41	1.41	
Error $_W$	26	186.45	7.17	

$^*p < .05$

The data reveal that no differences were occasioned by the projected visual and projected audio-visual modes of presenting test items. Item exposure time of a constant interval and a variable interval did not lead to significant differences between scores. When data were analyzed according to level of intelligence, a significant difference on the criterion variable was noted and it was expected. The only significant interaction occurs between mode of presentation and item exposure time. Because the half-tests were not of comparable difficulty, the interaction should be regarded as spurious. Unequivocal interpretations of the data are difficult to make owing to the lack of comparability of half-tests. Inspection of the data leads Es to believe that on items of constant exposure time the projected audio condition produces higher scores than does projected alone. Also, it seems for high-scoring Ss on the ACE the projected audio condition is superior to the projected condition alone on items of variable exposure time; however, this seeming difference does not hold for Ss having low ACE scores.

Summary

Groups of college freshmen and sophomores who were registered in an American Government course were administered a fifty-item multiple-choice test constructed by the course instructor. One group was administered the test under normal conditions. Another group took the test visually in the sense that each item, one item at a time, was projected onto a screen which served as the only exposure of that item to the S. The third group took the test under projected conditions, but in addition when the item appeared on the screen it was read to them by means of a tape recording. Under the projected conditions, a block of twenty-five items received a constant exposure interval and the remaining items received variable exposures in accordance with the reading load of each item. The order of blocks was different for the two groups; thus, one group had constant exposure intervals on items 1-25, whereas the other group had constant exposure intervals on items 26-50. The analysis of data revealed that no between group (one control and two experimental) differences occurred on either half-test or the whole test. However, analysis of the C group data revealed that the half-tests were not of comparable difficulty. Further analyses were confined to the two X groups. The analyses revealed no significant between group differences could be attributed to mode of presentation (projected, and projected audio) or to item exposure time (constant and variable). However, the interaction of these variables was significant but it was regarded as spurious due to the lack of comparably difficult half-tests. There was some evidence that the projected audio condition is superior to the projected condition alone on items of constant exposure time.

Generally, the data support the notion that the administration of tests by projecting them onto a screen to be viewed and responded to by students produces scores similar to those obtained under normal administrative conditions.

SOCIAL STUDIES[1]

Purpose

This study compares normal, projected-silent, and projected-audio modes of test administration. In addition, relationships between the data and how Ss were taught (televised instruction or the absence of it) and Ss' mental ability are investigated. A primary concern was with transfer of training from mode of instruction to mode of testing. It was hypothesized that Ss who had been taught by television would achieve greater scores when tested by projected-silent and projected-audio methods than would Ss who had not been taught by television.

Procedure

Subjects. Ss consisted of twenty-nine matched pairs of sixth-grade public school students. Pairs were matched on intelligence, a prognostic score which was a weighted combination of intelligence and achievement measures, and sex. Ss were drawn from each of three social studies classes in two schools. Ss from one school had had televised instruction in social studies for fifteen months prior to the study. In addition they received a televised enrichment hour and a reading hour each week. Their other school work was given under typical conditions in classes of normal size. The twenty-nine Ss from this school constituted the X group. The twenty-nine paired Ss, which constituted the C group, came from the other school and they had no experience with televised instruction during their school histories. Both the X and C groups received instruction over the same social studies content and for the same period of time. The schools were comparable with regard to socio-economic class and ability of students, educational facilities, and teacher effectiveness so far as it could be determined. Reading and mental ability ranges were the same for the X and C groups. Only twenty-nine matched pairs of Ss could be obtained. Data from all other students were discarded prior to data analysis.

Tests. The experimental tests were the Social Studies subtests of the Metropolitan Achievement Tests. One subtest appraises information and evokes verbal skills. The other subtest appraises study skills and evokes verbal and graphic skills. Forms A, B, and C of the subtests were used.

Testing Conditions. The experimental test was administered under all conditions to all students in each of the three classes in both schools. Form A was administered under normal conditions; Form B was administered under projected-silent conditions; and Form C was administered under projected-audio conditions. The effects of order of administration were controlled by assigning orders to classes using a 3 x 3 Latin Square basic design. Items were presented one at a time under projected conditions. Item order was according to order in the booklet conditions of the tests. There was no inter-item exposure interval during projected administrations. Each class was tested on three consecutive days, one

---

[1] Burr, William L. Empirical relationships among modes of testing, modes of instruction, and reading levels: in sixth-grade social studies. Unpublished doctoral thesis, School of Education, Florida State University, 1961. A condensation of this thesis, by the author, and bearing the same title as thesis, is in press with Journal Ex. Ed.

Form per day. The intra-item exposure interval on information items was fifteen seconds and the interval on study skills items was fifty seconds. These were constant exposure times. The intra-item exposure intervals were established on the basis of an earlier pilot study and their values are substantially less than they would have been had the allotted test taking time under normal conditions been prorated over all items in the respective subtests.

Analysis. Separate analyses of variance were run on the information subtest and the study skills subtest. The Newman-Keuls Sequential Range Test was used to make multiple non-significance tests on the differences between the ranked means of the significant interactions. A separate examination was made of the distribution of signed differences between projected-silent and projected-audio test scores and reading ability.

Results

Results of the analyses of variance of the information (verbal skill) subtest scores appear in Table 24. No differences could be attributed to main effects other than academic aptitude. Significant interactions were those of aptitude with method of presentation and the three way interaction of aptitude, method of instruction, and method of presentation. These differences may be attributed to academic grouping. The only significant trend was apparently the result of aptitude grouping, too.

Table 24

Analysis of Variance of Information Subtest (Verbal Skills)

Source of Variation	df	MS	*F*
Instructional Groups (G)	1	11.10	.12
Academic Aptitude (A)	2	8486.08	90.96*
G x A	2	85.58	.92
Error	100	93.29	
Test Presentations (T)	2	.005	.00
G x T	2	2.99	.14
A x T	4	73.30	3.48*
G x A x T	4	121.78	5.77*
Error	200	21.09	
Total	317		

*$p < .05$

Results of the analyses of variance for study skills (graphic and verbal skills) appear in Table 25. The only significant difference occurred between aptitude levels.

Table 25

Analysis of Variance of Study Skills Subtest (Graphic Skills)

Source of Variation	df	MS	*F*
Instructional Groups (G)	1	14.54	.13
Academic Aptitude (A)	2	6678.15	60.43*
G x A	2	287.51	2.60
Error	100	110.51	
Test Presentations (T)	2	.23	.00
G x T	2	16.36	.55
A x T	4	11.57	.39
G x A x T	4	54.29	1.84
Error	200	29.50	
Total	317		

*$p < .05$

The audio supplementation of the visual presentation of test items was analyzed further by relating discrepancies between projected-silent and projected-audio scores to reading comprehension scores. Scattergrams of reading scores and discrepancy scores suggest that for poor readers who have been taught by television, the addition of the audio stimulus increases their scores. However, good readers who had television instruction seem to be hampered by the audio stimulus during testing. Of the group who had received televised instruction, eight of the ten poorest readers scored higher during the projected-audio method whereas eight of the ten best readers scored lower during that method. This effect was not apparent for the C group nor did it appear on the study skills (graphic skills) subtest.

The results indicate that either projected-silent or projected-audio administration of tests produce results comparable to normal paper and pencil testing over the same content.

READING

Purpose

This study deals with determining whether changes in mean scores and the factor structure of a test occur if the test is administered under various conditions. The conditions are (a) normal, (b) projected-silent, and (c) projected-audio.

Procedure

__Subjects.__ __Ss__ were two groups of thirty each of white, ninth-grade public school students. Analysis of test data revealed that these groups were drawn from the same population.

Tests. All *Ss* were administered the School and College Ability Test (SCAT) and five tests of the Iowa Tests of Educational Development--Correctness and Appropriateness of Expression, Ability to do Quantitative Thinking, Ability to Interpret Reading Materials, Ability to Interpret Reading Materials in the Social Studies, and Ability to Interpret Reading Materials in the Natural Sciences. These were regarded as furnishing the independent or reference variables.

The dependent or criterion measures were scores from Forms Jm and Km of the advanced Reading Test (SRT) of the Stanford Achievement Test. This test consists of two parts, Part 1, Paragraph Meaning, and Part 2, Word Meaning.

Testing Conditions. All *Ss* were administered the reference tests and Form Jm of the Advanced Reading Test under normal conditions.

One X group was administered Form Km of the Advanced Reading Test under projected-silent conditions and the other X group was administered the same Form under projected-audio conditions. Common features of the experimental testings were; absence of an exposure interval between slides, item order conformed with order in the published test booklet, and total testing time was the same. In the majority of studies, one item was the exposure unit, but in this particular study, for Part 1, Paragraph Meaning, the exposure unit was a short reading passage on which one or more items was based. Consider the following sample item for Part 1, Paragraph Meaning.

> I am shorter than my sister and taller than my brother. This morning we stood beside one another. I looked down at my __51__ and __52__ at my sister.
>
> 51 1 friend 2 brother
> 3 sister 4 feet
>
> 52 5 around 6 back
> 7 up 8 down

The items in Part 2, Word Meaning are the usual multiple-choice vocabulary items. The exposure interval for each reading passage in Part 1 was proportionate to the reading load of the item as determined by Lorge's Readability Formula.[1] The exposure interval for Paragraph Meaning items ranged from 35-140 seconds. The exposure intervals for Word Meaning items were set at approximately sixteen seconds. The total exposure time over all items in a Form was equal to the testing time suggested by the test author for the booklet edition of the Form. The essential difference between the projected-audio and projected-silent administration was that in the former the reading passage and items were broadcast to *Ss* when the reading passage and items first appeared on the screen.

Analysis. Means, variances, within group intercorrelations, and *t* tests were computed. The intercorrelation matrix of reference variables were factor analyzed by the principal axis method and the resulting factors were rotated by the quartimax method. The loadings of the dependent variables on the rotated reference factors were estimated by a multiple regression technique.[1]

---

[1] Lorge, I. Predicting readability. *Teacher's College Record*, XLV (1944), 404-419.

Results

The analyses of means reveal that each projected condition on each test led to lower mean scores than the counterpart normally obtained mean scores. However, only on the Paragraph Meaning test were the differences significant at the .01 level. No differences were noted between means obtained by the two administrative conditions on either test. For the group which received the projected-silent condition, the correlations between their normal and experimentals are .62 on Paragraph Meaning, and .73 on Word Meaning. The comparable correlations for the group which took the projected audio condition were .61 and .88 respectively. These correlations suggest that normal and experimental scores on the Word Meaning test are equivalent.

Table 26

Means and Standard Deviations

	Group I, N = 28				Group II, N = 30			
Test	Paragraph Meaning		Word Meaning		Paragraph Meaning		Word Meanin	
Condition	P-S	N	P-S	N	P-A	N	P-A	N
Mean	35.7	37.6	36.8	37.1	34.9	36.7	38.0	38.4
S. D.	5.47	2.98	6.12	6.44	4.25	4.33	4.70	5.66

P-S = Projected silent
N = Normal
P-A = Projected audio

---

[1] Fruchter, B. Some factor analytic designs based on reference batteries. Paper read at APA symposium on "Selection of Variables for Factor Analysis," Chicago, Ill., September 7, 1960.

Table 27

$\underline{t}$ Test of Differences Between Means[1]

	N_{ap}	N_{aw}	XAP	XSP	XAW	XSW
N_{sp}[2]	.93		2.81**	2.29*		
N_{sw}		-.81			-.60	.34
N_{ap}			2.96*	.77		
N_{aw}					.75	.10
XAP				-1.38		
XAW						.33

[1] Within group $\underline{t}$'s are corrected for correlated performance.

[2] N_{sp} normal scores of group who subsequently underwent projected-silent, Paragraph Meaning, condition. The letters in the other column entries indicate; X, projected; A and S, audio and silent, respectively; and P and W, Paragraph Meaning and Word Meaning, respectively.

* $p < .05$

** $p < .01$

The factor analysis of the reference variables yielded three factors which were named General, Quantitative, and Scientific. The estimated loadings of the dependent variables on the reference variables appear in Table 28. The data reveal that the experimental testing conditions yield results whose loadings on the reference variables are substantially the same as the loadings on these variables by normally obtained reading test results.

Table 28

Estimated Factor Loadings of Stanford Reading Test Scores

			Factors		
Form	Part	Condition	General	Quantitative	Scientific
Jm	1	Normal	.72	.19	-.27
Jm	2	Normal	.77	.30	-.21
Km	1	Proj.-Audio	.71	.22	-.34
Km	2	Proj.-Audio	.68	.26	-.12
Km	1	Proj.-Silent	.88	.12	.00
Km	2	Proj.-Silent	.79	.18	-.20

However, the loadings of the results of the projected-silent administration were slightly higher on the General factor and slightly lower on the Quantitative and Scientific factors than were the loadings of the normal and projected-audio administrations.

ARITHMETIC I

Purpose

The purpose of this study was to determine the similarity in factor structures of scores from (a) an arithmetic computation test administered under normal conditions and (b) from a comparable test in an adapted form that required recognizing a correct solution and which was administered under projected conditions. On one experimental administration the intra-item exposure interval was constant and on the other the interval was variable.

This investigation is essentially a follow-up of a study reported in an earlier section which reveals that Ss encountered difficulty in responding to arithmetic computation items when administered by projected methods. Two major difficulties were (a) the awkwardness in copying from screen to scratch-sheet, comparing scratch-sheet results to screen to locate the keyed response, and then referring from screen to answer sheet in order to record a response; and (b) the possibility of Ss unwittingly permitting their solution times to run beyond the time when the item (and its responses) disappeared from the screen and a new item appeared. The study presented here attempts to resolve such difficulties by determining whether recognizing a correct or incorrect solution to an item is equivalent to computing correctly the response to that item.

Procedure

Subjects. Ss consisted of two groups of seventh-grade, white public school students. There were twenty-seven Ss in one group and twenty-eight in the other. The groups were intact classes from the same school. School policy prescribed random distribution of students to classes. No tests were computed to determine if classes were from the same population.

Tests. The control test consisted of thirty-seven of the forty-five items constituting the Arithmetic Computation Test, Form Bm, of the Metropolitan Achievement Test. Eight of the original items were discarded because they emphasized graph reading rather than arithmetic computation.

The experimental test consisted of thirty-seven of the forty-five items constituting Form Cm of the Metropolitan Achievement Test. The items were altered so that Ss could respond to them without computing responses. The general method of adaptation is shown below.

Suppose the original item was:

12. $9\overline{).081}$

A. 9
B. .9
C. .09
D. .009

Then it might have been adapted as follows:

12. $9\overline{).081}$

$$\begin{array}{r} .09 \\ 9\overline{).081} \\ \underline{81} \end{array}$$ <----ANSWER

S would respond 'Right' or 'Wrong' depending on his appraisal of the correctness of the given "Answer."

Eighteen of the items were supplied with correct computations and answers. The remaining nineteen were incorrect. Incorrect answers were always accompanied by incorrect computations when it was necessary to show the computations as would be the case in 7/8 + 9/16. Items were randomly assigned item numbers, thereby violating the order of items which appear in the comparable booklet edition.

Reference tests consisted of Verbal Meaning, Reasoning, Spatial Relations, and Number Reasoning from the Primary Mental Abilities Test; the SRA Clerical Speed and Accuracy Test; and the Writing Digits Test of the Kit of Selected Tests for Reference Aptitude and Achievement Factors.

Testing Conditions. One X group was administered the reference battery under normal conditions, the experimental arithmetic test under projected conditions with variable item exposure times, and the control arithmetic computation test under normal conditions, in that order. The other X group was administered the reference battery under normal conditions, the experimental arithmetic test under projected conditions with constant item exposure times, and the control arithmetic computation test, in that order.

Total test-taking time for the projected arithmetic test was set at exactly half (fifteen minutes) that of the control arithmetic test (thirty minutes). Past experience indicates that less testing time is necessary when tests are administered under projected conditions; furthermore, it was felt that the "recognition" test would require less time than would the "computation" test.

The constant item exposure interval was set at 24 seconds which is approximately equal to fifteen minutes divided by thirty-seven items. Variable exposure intervals were: twelve items at 12 seconds each; thirteen items at 24 seconds each; and twelve items at 36 seconds each. Each item had been sorted into one of three categories on the basis of E's judgment of the number of operations involved in solving the item and the apparent difficulty of the item, and on the basis of rather inaccurate records of time spent on each item by students who had participated in a pilot study.

During the testings, there was no inter-item exposure interval, items were presented one at a time, and item order was the same on both X testings but differed from item order in the booklet edition.

Analysis. Means, standard deviations, and correlation coefficients were computed for the X and C tests. The reference battery was factored by the principal axis method. The obtained factors were tested for significance and then rotated by the quartimax method. The loadings of the C and X tests on the reference factors were then estimated by a regression technique.

Results

Data bearing on means, standard deviations, and validity of the C and X tests appear in Table 29. The statistics indicate that means, on X and C computed on a Rights basis are comparable but variances on the X tests were substantially less than on the C tests. The "between forms" correlations, .67 and .61, are slightly less than that reported by the test author between the two forms when administered under normal conditions.

Table 29

Data on Control and Experimental Arithmetic Tests

Group	Exposure	Scoring	r_{xc}	Means		Standard Deviations	
				X	C[a]	X	C[a]
1	Variable	R-W	.69	22.3		7.2	
1	Variable	R	.67	29.2	30.5	4.1	6.6
2	Constant	R-W	.61	20.7		8.3	
2	Constant	R	.61	28.5	28.1	4.5	5.2

[a] Booklet edition does not suggest a correction for guessing. Thus for Group 1 on R-W scoring the correlation is between R-W on X and R on C.

The factor analysis of the reference tests resulted in six independent factors. Each of the reference variables loads highly

(.88--.97) on one and only one factor, and only one variable loads substantially on all factors. The data appear in Table 30.

Table 30

Rotated Factors

Tests	Factors I	II	III	IV	V	VI
V	.20	.08	.09	.23	.94	.09
S	.04	-.02	.14	.03	.18	.97
R	.88	.12	.24	.11	.34	.17
N	.01	.11	.10	.97	.19	.08
CSA	.15	.12	.94	.14	.08	.22
WD	-.08	.97	.17	.06	.10	.05

The projections of X test R-W scores on the reference factors appear in Table 31. It shows that the experimental conditions may not lead to the measurement of the same constructs as do normal test conditions. This possibility is supported by the differences in loadings on the CSA and Writing Digits factors. Differences between loadings on V, X, N, and CSA for the constant exposure are probably due to forced pacing on the longer or more difficult items. This pacing factor would not be as important under variable timing or normal administration.

Table 31

Projection of Arithmetic Tests (X and C) onto Reference Factors

Condition	Factors Verbal	Space	Reasoning	Number	CSA	Writing Digits
X-variable	.23	.20	.11	.37	.57	.21
Control	.23	.23	.17	.37	.22	.33
X-constant	.18	.13	.23	.16	.26	-.01
Control	.28	.23	.23	.31	.02	-.04

The difference in projected factor loadings of the two control tests on CSA and Writing Digits is accounted for by smaller correlation coefficients on those two tests by the group taking the constant time test. This could be the result of class differences in ability or discipline, or an administrative factor. E feels that this between group difference does not detract from the significance of the within group differences.

ARITHMETIC II

Problem

In the preceding study, some evidence is presented bearing on the seeming superiority of variable intra-item exposure intervals over constant intra-item exposure intervals when testing arithmetic computation by means of items which appraise S's ability to perceive correct solutions rather than to compute responses correctly. The study presented here deals with the extent of similarity of results obtained by administering a computation test in the usual way and the results obtained when comparable items are adapted to recognition items and are administered by projected methods. Comparisons are drawn with regard to means, variances, reliabilities, and factor loadings.

Procedure

Subjects. Ss consisted of 159 white males and females who were enrolled in the ninth-grade class of white, public schools in Florida.

Tests. Form E of the Arithmetic Computation Subtest of the Metropolitan Achievement Tests served as the control test and Form C, in an adapted form, served as the experimental test.

The adaptation of each item for the experimental test consisted of adding to it a worked-out solution and answer. For twenty-three of the items the appended solutions were correct; for twenty-two items, the solutions were incorrect. Incorrect solutions were based on common errors committed by students when attempting to solve such items.

The following tests were administered to obtain reference data, The Reading, Language, Arithmetic Problem Solving, Social Studies, and Science Subtests of the Metropolitan Achievement Tests; the Clerical Speed and Accuracy Test of the Differential Aptitude Tests; the Writing Digits test of the Kit of Selected Tests for Reference and Achievement Factors; and the School and College Ability Test.

Testing Conditions. The control Arithmetic Computation test, the other tests in the Metropolitan Achievement Test battery, and the School and College Ability Test were administered during one test session. At a subsequent session three weeks later the experimental form of the Arithmetic Computation Test, the Writing Digits Test, and the Clerical Speed and Accuracy Test were administered.

The experimental test was presented one item at a time. The total test-taking time for it was reduced to seventy per cent of the time allotted to its comparable form when administered normally. Intra-item exposure interval was variable: one-third of the items received twenty-four seconds; another third, thirty-six seconds; and the remaining third, forty-eight seconds. Items were categorized on the basis of a preliminary study executed to determine the mean solution time for each item. There was no inter-item exposure interval.

Directions were given by means of a precut magnetic tape which was broadcast immediately after testing materials were distributed.

Analyses. Means, standard deviations, and reliabilities were computed for the control and experimental administration. A factor matrix was computed for the experimental and control test scores and the scores yielded by the reference battery.

Results

The mean scores on the control and experimental tests did not differ significantly at the .05 level after the data were corrected for correlated performance. Variances were significantly different at the .05 level; the control variance was larger. The KR-20 reliabilities were high and not significantly different. These statistics appear in Table 32.

Table 32

Means, Standard Deviations, and Reliability Coefficients

	Control	Experimental	P
Mean	33.87	34.04	> .05
Standard Deviation	7.14	6.34	< .05
Reliability	.87	.83	> .05

Correlation coefficients were computed between the control and experimental tests and the tests in the reference battery. These correlations appear in Table 33. Control test results correlate higher with all tests in the reference battery except for Writing Digits and Clerical Speed and Accuracy than do experimental test results. The control and experimental test correlate at .74 which is of a lower magnitude than one would expect if the forms were truly equivalent. This suggests that the adaptation possibly reduced the equivalency of the form; i.e., recognizing the correctness of a computation is not equivalent to performing the computation which is called for.

Table 33

Correlation of Experimental and Control Tests with Reference Battery

Variable	Male		Female	
	Control	Experimental	Control	Experimental
Control	1.00	.74	1.00	.74
Experimental	.74	1.00	.74	1.00
SCAT Verbal	.64	.55	.69	.55
SCAT Quantitative	.80	.74	.85	.72
Reading	.65	.49	.70	.61
Language	.59	.54	.70	.62
Problem Solving	.83	.72	.88	.75
Social Studies	.64	.56	.71	.57
Science	.63	.53	.60	.60
CSA	.37	.44	.52	.49
WD	.22	.16	.19	.23

The intercorrelation matrix, for males and females separately, was factor analyzed. Factors were computed for all Eigen values greater than one. The obtained factors were rotated by the varimax method. Three factors emerged which were tentatively named: I, Verbal; II, Clerical; and III, Quantitative. The rotated factor matrices appear in Table 34.

The experimental administration of the adapted form apparently leads to purer measures as regards quantitative ability. The experimental test is loaded less heavily on the Verbal factor than is the control test. Despite the reduced time allotted for the experimental test, the experimental and control tests load about the same level on the Clerical Speed and Accuracy Test.

Table 34

Rotated Factor Matrix

Variable	Male			Female		
	I	II	III	I	II	III
Control	.46	.14	.78	.49	.11	.79
Experimental	.24	.16	.89	.29	.17	.86
SCAT Verbal	.89	.06	.33	.89	.10	.30
SCAT Quantitative	.42	.14	.77	.56	.18	.69
Reading	.89	.16	.26	.86	.07	.37
Language	.60	.00	.45	.59	.27	.50
Problem Solving	.61	.09	.69	.63	.15	.68
Social Studies	.79	.22	.36	.85	.12	.33
Science	.85	.17	.30	.85	.19	.27
CSA	.12	.79	.33	.17	.71	.45
MD	.14	.88	.02	.12	.92	.01
Control*	.44	.10	.80	.51	.11	.77
Experimental*	.23	.11	.90	.22	.13	.86

*Corrected for guessing

Thus it would appear that the combination of adapting the computation items to recognition items and then administering them by projected methods provides a purer measure of quantitative ability than do typical items administered in the normal way. The data suggest that the combination of adaptation and projection destroys the equivalency of forms; i.e., the testing procedures are not interchangeable.

V. INTEREST

In the vocational interest area, only one investigation of the applicability of the projected method of testing was undertaken. Since the goodness of vocational interest measures depends greatly on the empirically determined stability of the obtained measures, and since the duration of the grant under which the investigations were conducted was comparatively brief, additional studies in this area were not undertaken.

Problem

The problem was to determine the relationships between criterion measures and sets of rankings of occupational groups for *Ss* who were administered the same occupational interest inventory *under* four sets of conditions, one C and three X conditions. Specifically, there were two periods of test administration. During the first period, data were gathered on: (a) a criterion measure which required *Ss* to indicate preferences for particular occupations; (b) an occupational interest inventory administered in the usual way; (c) the inventory administered one item at a time under projected conditions; and (d) the inventory administered as in (c) above except with reduced intra-exposure intervals. During the second period of test administration which occurred ten months subsequent to the first period, *Ss* who had been tested earlier were readministered the inventory *under* conditions (a) and (c) above.

Procedure

Subjects. *Ss* consisted of all 91 eleventh grade students in a white *Florida high* school; however, initial data were incomplete for two students so they were discarded. Seventy of the original *Ss*, all who were enrolled in school at the time, constituted the group *from* which stability data were gathered ten months later.

Tests. Criterion measures (CM) were collected from all *Ss*. *Ss* were *presented* with descriptions of jobs falling in each of *the six* occupational categories included in the occupational inventory and they were asked to indicate their preferences among jobs by ranking them.

The Occupational Interest Inventory (OII), Advanced Form, was used for both X and C administrations.

Testing conditions. For the C administration during the first and *second testing period* the OII was administered in accordance with directions in the test manual. The administration of the criterion measure (CM) was conducted under usual test administration conditions.

In all X administrations of the OII, there was no inter-item exposure interval, the item order was identical with item order in the booklet edition, and only one item was presented per exposure. Intra-item exposure interval differed in two of the experimental administrations. In the XN administration the intra-item exposure interval was twelve seconds which corresponds to total test taking time for the booklet edition administered under normal conditions divided by the number of items. In the XS administration, which occurred ten months subsequent to the XN administration, the exposure interval was also twelve seconds. In the XP administration, the intra-item exposure interval was eight seconds, two-thirds of the "normal" exposure time.

During the first testing period *Ss* were administered CM, C1, XN, and XP, in that order. A rest period and two "word games" were used between administrations in an attempt to reduce response sets and specific memory and, generally, to provide a change of pace. The four administrations occurred during a four hour period. Not all *Ss* were tested during the XP administration. *Ss* were randomly divided into two groups. One of these groups consisting of forty-four *Ss* underwent the XP administration.

During the second testing period, *Ss* were administered XS and C2, in that order. A short break occurred between administrations.

Analyses. Coefficients of concordance were computed on (a) all *Ss* for the CM, C1, and XN administrations, and on (b) all *Ss* who took all of the following--CM, C1, XN, and XP. Rho coefficients for each *S* were computed on pairs of CM, C1, XN, and XP. Various Signs Tests were computed to determine relative effectiveness of X administrations. Rho coefficients for each *S* were computed between (C1) (C2), (C1) (XS), and (XN) (XS).

Results

A coefficient of concordance (W) was computed on the ranks for each of the 89 *Ss* who completed the criterion scale and the C1 and XN administrations of the OII. The distribution of these values appears in Table 35. Note that *Ss* who subsequently took the XP administration appear separately in the frequency column appearing in the table. Each coefficient was tested for significance by the method outlined by Siegel.[1]

---

[1] Siegel, Sidney, *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill, 1956. Chapter 9.

Table 35

Distributions of Coefficients of Concordance for *Ss* on CM, and C1 and XP Administrations of OII

W Interval[a]	Frequency[b]	Frequency[c]	Total Frequency
.95 - .99	6	10	16
.90 - .94	13	7	20
.85 - .89	12	6	18
.80 - .84	4	11	15
.75 - .79	1	4	5
.70 - .74	3	4	7
.65 - .69	1	0	1
.60 - .64	1	1	2
.55 - .59	1	0	1
.50 - .54	1	1	2
.45 - .49	0	0	0
.40 - .44	1	0	1
.35 - .39	0	0	0
.30 - .34	0	1	1

[a] $W \geq .65$ is significant at .05 level

[b] *Ss* who took XP administration subsequently

[c] *Ss* who did not take XP administration

A coefficient of concordance was computed on the ranks for each of the forty-four *Ss* who, at the first session, completed the CM and all three administrations of the OII. The distribution of *W*'s, each based on four sets of ranks, appears in Table 36.

Table 36

Distributions of Coefficients of Concordance for Ss Who Completed CM and Cl, IN, and XP Administrations of OII

W Interval[a]	Frequency
.95 - .99	4
.90 - .94	12
.85 - .89	10
.80 - .84	7
.75 - .79	5
.70 - .74	0
.65 - .69	1
.60 - .64	1
.55 - .59	2
.50 - .54	1
.45 - .49	0
.40 - .44	1

[a] $W \geq .55$ significant at .05 level

Rank order coefficients were computed between pairs of ranks on CM, Cl, IN and XP for each S. The distributions of these coefficients appear in Table 37.

Table 37

Distributions of Rho Coefficients Between Pairs of Ranks of Combinations of CM, Cl, XN, and XP

	Rho Coefficients					
Rho Interval	(CM)(Cl)	(CM)(XN)	(CM)(XP)	(Cl)(XN)	(Cl)(XP)	(XN)(XP)
.96 - 1.00	10	6	4	40	12	24
.91 - .95	11	14	5	16	11	5
.86 - .90	11	13	6	11	8	2
.81 - .85	16	13	4	3	4	3
.76 - .80	4	7	6	3	4	3
.71 - .75	9	9	2	2		
.66 - .70	3	3	1	1	1	1
.61 - .65	5	6	4	3	2	2
.56 - .60	7	8	4	2	1	
.51 - .55	3	1	1	2	1	
.46 - .50	5	3	2		1	1
.41 - .45	1		1	1		2
.36 - .40	1	2	1	1		
.31 - .35			1			
.26 - .30						
.21 - .25	1	1	1	1		
.16 - .20	1	2	1			
below .16	1	1		3	1	2

2

Sign tests were computed between rank correlations for various combinations of the four obtained measures. A summary of these comparisons appears in Table 38.

Table 38

Sign Tests for Selected Comparisons

Comparison		Number	Differences	Probability
CM vs. Cl, with CM vs. XN	$r_{(CM)(XN)} > r_{(CM)(Cl)}$	42		
	$r_{(CM)(Cl)} > r_{(CM)(XN)}$	38	4	.73
	$r_{(CM)(Cl)} = r_{(CM)(XN)}$	9		
CM vs. Cl, with CM vs. XP	$r_{(CM)(Cl)} > r_{(CM)(XP)}$	24		
	$r_{(CM)(XP)} > r_{(CM)(Cl)}$	17	7	.35
	$r_{(CM)(XP)} = r_{(CM)(Cl)}$	3		
Cl vs. XN, with C vs. XP	$r_{(Cl)(XN)} > r_{(Cl)(XP)}$	22		
	$r_{(Cl)(XP)} > r_{(Cl)(XN)}$	13	9	.18
	$r_{(Cl)(XP)} = r_{(Cl)(XN)}$	9		

Rank order coefficients indicating stability were computed for the *Ss* still in school at the second testing session. Distributions of these coefficients appear in Table 39.

Table 39

Distributions of Rank Coefficients Indicating Stability

Rho Interval	(C1)(C2)	(C1)(XS)	(XN)(XS)
.96 - 1.00	18	15	17
.91 - .95	9	12	13
.86 - .90	8	6	13
.81 - .85	11	11	5
.76 - .80	2	9	6
.71 - .75	5	4	1
.66 - .70	2	2	2
.61 - .65	3	0	3
.56 - .60	1	1	3
.51 - .55	2	1	1
.46 - .50	3	2	1
.41 - .45	1	1	2
.36 - .40	1	1	
.31 - .35		1	1
.26 - .30	1		
.21 - .25	1		
.16 - .20			
below .16	2	4	2

The data in Table 35 reveal that 78 per cent of the W's computed on rank orders of occupations given by each S on the CM, C1, and XN administrations were equal to or greater than .80; and that 92 per cent of the W's were significant at least at the .05 level. The distribution of W's in Table 36, which were calculated on CM and the C1, XN and XP administration of the OII, reveal that 75 per cent of them were equal to or greater than .80; and that 95 per cent were significant at least at the .05 level of significance. Thus, one might generalize that high relationships exist among the sets of occupational ranks which Ss produced.

The data in Table 37, which presents distributions of rho coefficients between pairs of ranks obtained under different administrative conditions, reveal that the following percentages of rho coefficients exceeded .80; between CM and Cl, 54 per cent; between CM and XN, 52 per cent; between CM and XP, 43 per cent; between Cl and XN, 79 per cent; between Cl and XP, 79 per cent; and between XN and XP, 77 per cent. Thus, it seems that CM rankings relate less strongly to the sets of rankings obtained by various administrations of the OII than the latter sets relate to each other. Thus, one might conclude that rankings obtained by use of CM did not serve as good predictions of rankings obtained on the OII which, due to its length and prior validation, was surely more reliable and valid. Consequently, one must assume that the W's which appear in Tables 35 and 36 underestimate the strength of relationship between various administrations of the OII owing to the introduction of the somewhat irrelevant CM rankings. The percentages of rho's that exceed .80 suggest that the Cl, XN, and XP administrations of the OII are somewhat interchangeable.

The most pertinent sign test is the comparison of (Cl) (XN) with (Cl) (XP) which, although not significant, reveals that the coefficients between (Cl) (XP) are generally lower than those between (Cl) (XN). Inasmuch as the Cl administration corresponds to the administration of the OII under normal conditions, it appears that of the two projected administrations of the OII the XN administration is superior. The item exposure interval on the XN administration was the same as that on the Cl administration if all Ss used all the time available for the Cl administration and divided their time equally over all items. The XP administration, in which item exposure time was reduced 33 per cent from the XN administration, yields rankings which generally relate less strongly to Cl rankings for each S than do XN rankings.

The data in Table 39, which presents the distributions of the rho coefficients obtained under the same administrative conditions with a time lapse of 10 months, reveal that 50 per cent of the coefficients for Cl and C2 exceed .85 and 61 per cent of the coefficients for XN and XS exceed .85. In addition, 13 per cent of the coefficients for Cl and C2 are .50 or lower, whereas 9 per cent of those for XN and XS are below this limit. Thus it appears that the stability of the rankings obtained under the projected mode of administration is at least as good as that obtained under normal administrative conditions.

Summary

The purpose of this study was to determine the relationships obtained by Ss who were administered the Occupational Interest Inventory under normal administration conditions and two experimental conditions which involved administering the OII by projecting items one at a time onto a 3' x 5' screen. The item exposure time on one experimental administration was equal to the time allotted for the OII when administered under normal conditions divided by the total number of items. The item exposure time on the other experimental administration was two-thirds that of the former and it was regarded as the "paced" administration. In addition Ss were asked to complete a criterion scale which involved their ranking, according to their preferences, the occupations for which scores are obtained on the OII. The data revealed that the criterion measure was relatively invalid. Both the normal administration and the projected normal administrations were repeated for students still in school ten months later.

Coefficients of concordance were computed on occupational rankings yielded on the criterion measure, and the control and normally-timed experimental administrations; a separate set of coefficients was computed on the above sets of rankings and the rankings from the experimental-paced administration. The distribution of W's revealed that at least three-quarters of them were equal to or greater than .80, and approximately 90 per cent of all W's were significant at least at the .05 level of confidence. The W's were regarded as underestimates of the relationships between OII rankings obtained under various conditions because the rankings yielded by the criterion measure entered into the calculation of W coefficients and it was subsequently learned that rankings on the criterion measure were not highly associated with OII rankings. Rho coefficients between pairs of rankings revealed that 75 to 80 per cent of all coefficients computed between pairs of control and the two experimental rankings were equal to or greater than .80. A sign test revealed that rho coefficients between the control and experimental normally-timed rankings were generally greater, but not significantly so, than the coefficients between the control and experimental paced rankings on the OII. Rho coefficients were computed to estimate the stability of rankings, over a ten-month period, on the normal and projected normal administration. Analyses revealed that the distributions of coefficients did not differ significantly.

The analysis lends support to the belief that vocational interest inventories such as the OII yield similar sets of rankings when the items are administered by projecting them on a screen as when the items are presented in booklet form. Furthermore no differences were noted between the distributions of stability coefficients of control and experimentally obtained rankings.

VI. ADJUSTMENT AND PERSONALITY

Three studies are reported in this section. Two deal with the administration of a delinquency-proneness scale under normal and paced-projected condition. The second of these two is essentially a cross-validation of the first; however, in some respects the initial study is extended in the second. The third study reports an investigation of the similarity of scores obtained when the Minnesota Multiphasic Personality Inventory is administered under normal and projected conditions.

Typical tests of adjustment and personality which are frequently administered in the public schools were ignored for one or both of two reasons. First, current sentiment about the propriety of administering such tests to public school students was such that it seemed sensible to avoid them for the time being. Second, the absence of, or extreme difficulty in obtaining valid criterion data for such instruments, would make it difficult to appraise the relative concurrent validity of sets of scores obtained under normal and projected conditions.

DELINQUENCY-PRONENESS: I[1]

Problem

When one considers pacing Ss through a personality or adjustment inventory, one may hypothesize that the paced-projected testing would elicit different response patterns from Ss from those they would give if the same inventory were administered to them under normal conditions. First, paced-projected testing requires that the items be answered quickly; thus, a response is more likely to be a first impression or snap reaction instead of one which is carefully, reflectively, and leisurely thought out. The directions for administering the California Test of Personality, the S-O Rorschach Test, and the Minnesota Personality Scale request the examinee to respond rapidly on the assumption that this first impression to items in these scales is a valid and representative expression of his true feelings. Second, if only one item is shown at a time and only once during the test, S is not able to respond in terms of subsequent items nor is he likely to respond in terms of previous items. Consequently, his opportunity to falsify, to create an impression, or to assume a response set is reduced.

It was hypothesized (a) that a sample of delinquent boys would attain a higher mean score on a delinquency-proneness scale administered to them under paced-projected conditions than would a comparable group of delinquents who were administered the same scale under normal conditions; (b) that two comparable groups of nondelinquents would manifest the same order of mean scores as that described in (a) above; and (c) that the four obtained means would be ordered as follows: delinquent paced-projected > delinquent normal ≥ nondelinquent paced-projected > nondelinquent normal.

---

[1] King, F.J., Kropp, R.P., Crawford, W.R., and Moyel, I.S. "Changes in delinquency-proneness scores associated with paced-projected and normal modes of test administration." Psychological Reports. Vol. 9 (1961), pp. 55-58.

Procedure

Sample. The delinquent group consisted of sixty-four *Ss* who were randomly selected from the population of an industrial school for boys. The group was drawn from that part of the population that met each of the following criteria: (a) in residence at least six weeks, (b) reading level equal to or greater than fifth-grade level, (c) at least one criminal offense on record, and (d) between 15 and 16 years of age. This group was then subdivided randomly into two groups.

The nondelinquent group consisted of 32 eleventh-grade boys enrolled in a white public school who met criteria (b) and (d) above. This group was then subdivided randomly into two sub-groups.

Tests. Three scales, designed to identify juvenile delinquents, which were developed by Quay and Peterson[1], Gough and Peterson[2], and Hathaway and Monachesi[3], were slightly edited and combined into one scale. Each of the three scales consisted of true-false items that had been evaluated by their respective authors against delinquent and nondelinquent populations. The investigators eliminated some items from these scales on the basis of item content that, it was felt, might dispose school officials not to permit students for whom they were responsible to participate in the study. The resulting scale was named "High School Opinion Survey," hereafter referred to as HSOS, to disguise its true nature from *Ss*. It contained 92 true-false items and, theoretically, the higher the *obtained* score, the greater the delinquency-proneness.

The Iowa Silent Reading Test, New Edition, Adv., Form CM, was administered initially to all potential *Ss* to estimate their reading rate and comprehension. It was felt *that* all *Ss* should be able to read at least at the fifth-grade level so *that they* would be able to understand the HSOS items.

Method of testing. One delinquent and one nondelinquent group were *administered the* HSOS in the normal paper-and-pencil manner. The other delinquent and nondelinquent groups were administered the HSOS under the paced-projected method. Each item was projected individually for 6 sec. during which Ss read and responded to it. There was no inter-exposure interval. *All* four groups recorded their responses on standard IBM answer sheets.

Method of analysis. The data were analyzed according to a 2 x 2 x 2 *factorial design:* delinquent and nondelinquent, projected and regular test administration, and high and low reading ability.

Results

Results of the analyses appear in Table 40. Mean scores for the eight sub-groups appear in Table 41.

---

[1] Quay, H., and Peterson, D.R. A brief scale for juvenile delinquency. *J. clin. Psychol.*, 1958, 14, 139-142.

[2] Gough, H.G., and Peterson, D.R. The identification and measurement of pre-dispositional factors in crime and delinquency. *J. consult. Psychol.*, 1952, 16, 207-212.

[3] Hathaway, S.R., and Monachesi, E.D. The personalities of predelinquent boys. *J. crim. Law, Criminol., Police Sci.*, 1957, 48, 149-163.

Table 40

Results of Analysis of Variance

Source of Variation		df	SS	MS	F
Delinquent vs. Non-Delinquent (D)		1	630.75	630.75	6.78*
Projected vs. Paper and Pencil (P)		1	495.04	495.04	5.32*
High vs. Low Reading (C)		1	1472.67	1472.67	15.82**
Interactions	DP	1	500.52	500.52	5.38*
	DC	1	402.52	402.52	4.32*
	PC	1	48.16	48.16	.52*
	DPC	1	1.33	1.33	.01*
Error		88	8190.64	93.07	
Total		95	11741.63		

$^{*}p < .05$

$^{**}p < .01$

Table 41

Sub-Group Means

	Delinquent		Non-Delinquent	
	High Comp.	Low Comp.	High Comp.	Low Comp.
Projected	44.56	48.25	39.87	51.75
Paper and Pencil	42.00	48.18	27.12	42.50

The HSOS discriminated significantly between the delinquent and nondelinquent groups regardless of test administration. The paced-projected presentation produced a significant increase in scores regardless of groups. The significant interaction between groups and treatments indicated the shift in scores was more pronounced for the nondelinquents than for the delinquents. Delinquency-proneness scores were significantly higher for Ss having low reading ability than for Ss having high reading ability thus supporting an earlier finding by Quay and Peterson. The hypothesis about the order of the four means was **sustained** when level of comprehension was disregarded. When comprehension, group membership, and method of presentation are all considered, the highest mean score was obtained by the nondelinquent, low comprehension, projected group. However, it is doubtful whether this finding has any real meaning inasmuch as the second order interaction (DPC) failed to reach significance at the .05 level.

Results of this study might be explained, at least tentatively, at several levels. The paragraph above is confined largely to mean differences that are associated with different methods of test administration. Possibly the results might be explained very hesitantly or "rationalized" by resorting to a more fundamental level. Perhaps both the delinquent and nondelinquent groups are characterized by similar levels of delinquency-proneness; however, it is manifest for the former and latent for the latter. In turn might it be manifest for one and latent for the other because non-delinquents (latent) refrain from acting until they evaluate possible actions against social norms whereas delinquents (manifest) act without doing so. In a small way, the results support this notion inasmuch as nondelinquents receive considerably higher delinquency-proneness scores when they are not permitted to reflect leisurely about how to respond to items than when they do have such time available, whereas delinquents do not reveal the same magnitude of score differences.

DELINQUENCY-PRONENESS: II

In the study presented above, it was found that the paced-projected administration produced significantly higher scores than did the normal mode regardless of the classification of the individual, but the increase in mean score was greater for normals than for delinquents.

Problem

The study reported here explores further the effect of paced-projected methods of administration on the performance of normal subjects in responding to delinquency-proneness items. Under consideration were three levels of intra-item exposure interval, 3, 6, and 9 seconds. Also of interest were possible sex differences in scores on the instrument and interactions of these differences with methods of administration.

Procedure

Subjects. Ss were approximately 300 male and female students from two white public high schools. The Ss were drawn from twelfth-grade English classes.

Test. The delinquency-proneness questionnaire was the same as that used in the previous study. The paper and pencil form was called the High School Opinion Survey (HSOS) and the projected form, which consisted of the same items rearranged into a different order, was called the Projected Opinion Survey (POS).

Testing Conditions. Ss responded to both tests, one half taking the HSOS first and the other half taking the POS first. During the interval between administration of the delinquency items, Ss took a fifty-item paper-and-pencil test containing six scales dealing with aggression, impulsion, conservatism, displaced oral aggression, deliberation, and passivity. These data are not germane to this study, but will be used in another study which will incorporate a part of the present study with data yet to be collected. At the time of presentation of the POS, each group was divided into three subgroups consisting of members of both sexes, and each subgroup was administered the items under one of the three intra-item exposure times.

Analysis. The testing arrangement allowed the data to be subjected to a 3 x 2 x 2 x 2 analysis of variance design (three exposure times, two sexes, two order of presentation, and two test.) Ss were randomly eliminated from twenty-three of the subgroups until each contained nineteen Ss, the number of Ss in the twenty-fourth group. Since each S served as his own control through taking both HSOS and POS, it was not considered necessary to take into account other variables that might influence the way S responded to the items. However, it was felt to be necessary to include in the analysis the order of presentation of tests since it is possible that one testing experience would influence later responses to the same items.

The Newman-Keuls Sequential Range Test was used to test differences between individual means in significant effects. The scores which were analyzed were the proportions of keyed responses to total number of items responded to, thus ignoring differences in omitted items.

Results

The results of the analysis appear in Table 42. Mean scores for the highest order interaction that exceeded the .05 level (IOT) are given in Table 43. The means for the other significant effects, except sex, are contained within this interaction. The mean score for males was 31.40, for females, 24.46.

When applied to the IOT interaction, the Newman-Keuls Test indicated that control test (HSOS) differences existed both within order one and between the two orders of administration. While it is possible to discuss POS differences in the second order without regard to HSOS scores, the remainder of the interaction must be considered in terms of the HSOS-POS difference for each of its subgroups. When HSOS was administered first (order one), there were no HSOS-POS differences for any of the three intervals. Under the second order, significant differences in HSOS-POS scores did appear for both the 3" and 9" intervals. In addition all POS scores were significantly different with 3" > 9" > 6". Apparently the order influence in this interaction as well as the significant order main effect can be explained by the higher scores produced by the paced-projected condition but modified by the tendency of Ss to answer items the second time in the same way as they answered first. The interaction IOT contains the other significant effects, T, IO, IT, and OT, and the explanation applied to it probably pertains to them.

Table 42

Results of Analysis of Variance

Source of Variation	df	MS	F
Variation between S's	227		
Interval (I)	2	228.49	1.61
Sex (S)	1	5502.32	38.81**
Order (O)	1	3296.22	23.25**
Interaction IS	2	48.50	.34
IO	2	986.17	6.96**
SO	1	116.01	.82
ISO	2	48.91	.34
Error_B	216	141.77	
Test (T)	1	526.54	31.06**
Interaction IT	2	497.90	29.37**
ST	1	2.53	.14
OT	1	424.56	25.05**
IST	2	28.96	1.71
IOT	2	486.56	28.71**
SOT	1	16.98	1.00
ISOT	2	4.67	.28
Error_W	216	16.95	
Total	455		

$^{**}p < .01$

Table 43

Means for IOT Interaction

	HSOS First			HSOS Last		
	3"	6"	9"	3"	6"	9"
HSOS	23.97	27.42	24.00	28.31	29.00	28.53
POS	25.08	28.63	22.34	39.53	26.32	32.13

In general, the results tend to support the conclusions of the former study that paced-projected administration of the delinquency-proneness scale causes normal Ss to attain higher, i.e., more delinquent-prone, scores than does the normal administration. One inconsistency exists in the present findings in that the 9" interval produced higher mean scores than 6". The explanation of this difficulty is not readily apparent. In addition, it should be noted that the over all mean score for non-delinquent boys in this study was considerably lower than the mean score for non-delinquent boys in the former study (31.40 and 40.31). No explanation of this finding is presently possible.

Since the effect of the paced-projected procedure found in the first study is regarded as corroborated by this study, it is felt that future research might be directed toward determining whether the scores produced by the paced-projected method are more valid, i.e., more predictive of delinquency, than are scores produced under typical testing conditions.

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

Problem

This study investigates differences between scores obtained on the Minnesota Multiphasic Personality Inventory (MMPI) under the paced-projected method of administration and scores obtained under the normal pencil and paper administration.

The first part of the study deals with whether or not gross differences in personality measures are observed between the two methods of administration.

The second part of this study deals with changes in scores on the validity scales of the MMPI obtained under the two methods of administration. It has been assumed traditionally that capturing S's first impressions of items wold lead to more valid data than are obtained when S takes time to reflect on response acceptability and has the opportunity to compare his responses to those of previous and subsequent items. Consequently, his opportunity to falsify responses in order to place himself in a good, or perhaps bad, light is reduced. Responses to the L and K scales reflect S's tendency to be over-critical of himself or to place himself in a socially acceptable light.

On the other hand, the paced-projected technique might lead to *S*'s being unable to reread items which he fails to comprehend fully or his being unable to completely read the longer items. The former would lead to random marking or to errors in responding. The latter would lead to random marking or the omission of items. Changes in the "?" and the F scale measure differences in validity due to these factors.

Procedure

Subjects. The sample for the first part of the study consisted of two classes of undergraduate students. One class consisted of 17 males and 40 females; the other class consisted of 23 males and 20 females. The *Ss* were divided into four groups; i.e., two classes x two sexes. The sample for the second part of the study consisted of 14 *Ss* chosen randomly from the *Ss* in each of the above four groups who had a complete set of data.

Tests. The first 366 items of the MMPI were administered both normally and experimentally to all *Ss*.

The clinical scales used in the first part of this study are as follows:

Hs - Hypochondriasis
D - Depression
Hy - Hysteria
Pd - Psychopathic Deviate
Mf - Masculinity-Femininity
Pa - Paranoia
Pt - Psychasthenia
Sc - Schizophrenia
Ma - Hypomania

The four control scales of the MMPI were also treated in the first part of the study. They are described subsequently.

The control scales of the MMPI which are used as experimental variables in the second part of the study are described as follows:[1]

1. "?" Scale. The "?" score consists of the number of omissions. Large "?" scores invalidate the other scales since scores are interpreted in terms of the number of keyed responses. Profiles are considered invalid if more than one out of fifteen items are omitted.

2. F Scale. The F scale consists of 64 items which are answered with a relatively low frequency in either the true or false direction. On each item the scored response is that response which is rarely made by unselected normal subjects. The items were chosen to include a wide variety of content so that it would be unlikely that a particular pattern of response would lead to an individual's answering many items in the unusual direction. The mean number of these items which are answered in the unusual way runs from two to four for groups of normal subjects.

---

[1] Meehl, P.E. and Hathaway, S.R. "The K factor as a suppressor variable in the MMPI." in Welsh, G.S., and Dahlstrom, W.G. *Basic Readings on the MMPI*. Minneapolis: Un. of Minn. Press, 1956.

The F scale is interpreted as indicating that the subject made mistakes in recording his responses or that he did not understand the items adequately enough to give conventional responses. High F scores should indicate records of subjects that could not comprehend the items or who failed to cooperate sufficiently to give expected responses.

3. L Scale. The 15 items which constitute the L scale present situations which are desirable socially but which are rarely true of an individual. Normal groups have means ranging from three to five on these items, very few individuals score above seven. High L scores are interpreted as indicating intent to deceive because the subject places himself in an unusually good light.

4. K Scale. The K scale consists of 30 items which were empirically chosen as measures of test-taking attitude. The content of these items suggest an attitude of denying clinically unhealthy symptoms together with a disposition to see only good in others as well as oneself. Relatively low K scores might indicate defensiveness leading to under-interpreting otherwise deviant profiles whereas relatively high K scores indicate "plus-getting;" i.e., a tendency to be overly honest and self-critical.

Testing Conditions. One classroom group was administered the MMPI in the normal paper-and-pencil manner. On the same day the other classroom group was administered the MMPI by the paced-projected method. After an interval of two days, Ss were retested; the former group received the paced-projected administration and the latter group received the normal paper-and-pencil administration.

During the paced-projected administration, items were projected one at time. The intra-item exposure was five seconds and it was constant for all items. There was no inter-item exposure interval.

Analysis. The data gathered on the first day of testing were analyzed by dispersion analysis[1]. The data were treated in the following four independent groups:

1. Males receiving paced-projected administration (MX)

2. Females receiving paced-projected administration (FX)

3. Males receiving normal administration (MC)

4. Females receiving normal administration (FC)

The analytic technique examines the relationship of the four groups to all 13 variable; i.e., the nine personality measures and four control measures. The differences between the four groups were estimated by calculating the generalized distances between all six pairs of the four groups. The process involves the computation of a within groups cross-product matrix defined as

---

[1] Anderson, H. and Fruchter, B. Statistical Procedures: Multivariate Analysis with the Generalized Distance Function and Canonical Variates, Research Guide no. 3, Psychometric Laboratory, Department of Educational Psychology, University of Texas, May, 1957.

$$W_{ih} = \sum_{j=1}^{q} \sum_{r=1}^{n_j} (X_{ijr} - \bar{X}_{ij\cdot})\ (X_{hjr} - \bar{X}_{hj\cdot})$$

where q is the number of groups

n_j is the number of subjects in the j^{th} group

X_{ijr} is the score for the r^{th} subject in the j^{th} group on variable i

$\bar{X}_{ij\cdot}$ is the mean of the i^{th} variable for the j^{th} group.

W_{ij} was divided by $W = \sum_{j=1}^{q} n_j - q$ to form the dispersion matrix A_{ih}. A difference matrix d_i was formed for each pair of groups by obtaining the differences between each pair of means. The generalized distance between each pair of groups was computed by D^2 where $D^2 = d_i' A^{ih} d_i$ where d_i' is the transpose of d_i and A^{ih} is the inverse of A_{ih}. Each D^2 was tested by the F distribution where $\underline{F} = \frac{n_1 n_2}{n_1 + n_2}\ \frac{1}{W}\ \frac{W-p+1}{p}\ D^2$ where n_1 and n_2 are the sample sizes of the two groups considered, p is the number of variables, and the other terms are as defined previously. The *F* ratios obtained were evaluated at $p=13$ and $W-p+1=84$ degrees of freedom.

The second part of the study dealt solely with the four MMPI validity scales. Data from both testing sessions were used thus permitting *S* to serve as his own control. Data were analyzed by analysis of variance using the following variables as sources of variation: sex, order of testing (projected first or second), and condition (projected versus normal).

Results

The matrix of means for the four independent groups (data from the first testing session only) appear in Table 44. The generalized distances between groups and the significance tests of the generalized distances appear in Table 45.

All comparisons except (MX-MC) and (FX-FC) reflect across sex comparisons, hence these comparisons yield highly significant group differences. There appears to be no difference between the projected and normal administrations of the MMPI for males, but there is a significant difference between projected and normal scores for females.

The experimental administration led to a lower score than the normal condition on all clinical scales except Hy for females. However, for females all validity scores were higher, i.e., indicated less validity, for the experimental condition than for the normal condition.

The pattern for males was for scores to be lower for the experimental conditions. The exceptions were Hs, Ma, and "?".

Table 44

Means of Independent Groups on All Scales

Scale	Group			
	MX	MC	FX	FC
"?"	2.78	1.71	6.40	4.08
L	3.13	3.47	3.65	2.95
K	10.48	11.76	11.00	9.93
F	4.00	4.76	3.90	3.80
Hs	4.91	4.23	6.15	6.68
D	17.43	20.82	17.65	20.00
Hy	19.52	21.12	24.10	22.45
Pd	16.61	16.82	14.75	15.73
Mf	25.43	26.82	36.65	38.10
Pa	8.75	10.18	9.15	9.88
Pt	10.43	10.65	10.55	13.78
Sc	10.04	10.53	8.90	11.93
Ma	18.17	16.82	14.65	17.15

Whether or not the FX-FC differences can be attributed to the validity scale differences is the subject of the second part of the analysis.

An analysis of variance was computed for each validity scale. These analyses are based on repeated measures of 14 Ss in each of the four groups. Results of the analyses appear in Tables 46, F scale; 47, K scale; 48, "?" scale; and 50, L scale.

The analysis of the F scale indicates very little difference in F scores throughout the data. Apparently the experimental condition does not give rise to increased random marking and does not give rise to a greater misunderstanding of item content.

Table 45

Generalized Distances Between Groups and F-Tests of Group Differences

Groups Compared	Generalized Distances	F
MX - MC	1.45	.95
MX - FX	14.04	10.11^{**}
MX - FC	11.56	13.36^{**}
FX - MC	13.71	8.48^{**}
FC - MC	11.00	8.83^{**}
FX - FC	2.36	2.12^{*}

$^{*}p < .05$

$^{**}p < .01$

Table 46

Analysis of Variance of F Scale

Source of Variation	df	Mean Square	F
Sex	1	6.51	0.30
Order	1	2.58	0.12
Sex x Order	1	31.08	1.44
Error_B	52	21.62	
Condition	1	0.22	0.03
Sex x Condition	1	13.58	1.71
Order x Condition	1	0.22	0.03
Sex x Order x Condition	1	3.22	0.41
Error_W	52	7.93	

The analysis of the K scale indicates very little change in K scores across the data. It seems that neither test administration condition produces a greater response set than the other in terms of systematic bias due to test-taking attitude. There is no evidence that tendencies toward defensiveness or toward self-criticism are enhanced by either method.

Table 47

Analysis of Variance of K Scale

Source of Variation	df	Mean Square	F
Sex	1	6.51	0.30
Order	1	0.22	0.01
Sex x Order	1	47.58	2.22
Error$_B$	52	21.43	
Condition	1	0.44	0.15
Sex x Condition	1	1.51	0.51
Order x Condition	1	3.22	1.10
Sex x Order x Condition	1	0.72	0.25
Error$_W$	52	2.93	

Table 48

Analysis of Variance of "?" Scale

Source of Variation	df	Mean Square	F
Sex	1	295.75	2.07
Order	1	54.32	0.38
Sex x Order	1	132.89	0.93
Error$_B$	52	143.21	
Condition	1	120.14	3.89
Sex x Condition	1	20.57	0.67
Order x Condition	1	155.57	5.04*
Sex x Order x Condition	1	1.29	0.04
Error$_W$	52	30.89	

*$p < .05$

The analysis of the "?" scale reveals a significant interaction between order and condition. Also, since an F of 4.02 is required for significance, there is evidence of a mean difference between conditions. The means for the order by condition interaction are:

Experimental administered first	5.00
Experimental administered second	4.04
Normal administered second	0.57
Normal administered first	4.32

It appears that when the control test is administered first, Ss tend to omit the same number of items on the experimental test as they do on the normal test. But when the experimental test is administered first, Ss tend to omit fewer items when the test is given normally. Mean differences for this interaction are displayed in Table 49. These differences were tested for significance by multiple tests. The only significant difference occurs between the experimental test given first and the normal test given second.

Table 49

Mean Differences for the Interaction of Order With Condition on the "?" Scale

	X-2	C-1	C-2
X-1 Experimental given first	.96	.68	4.43*
X-2 Experimental given second		-.28	3.47
C-1 Normal given first			3.75
C-2 Normal given second			

$^*p < .05$

The difference between the means of the experimental and normal methods was almost significant (.06 > p > .05). The mean for the experimental administration was 4.52 omissions whereas the mean for the normal group was 2.45 omissions. This is another reflection of the interaction previously discussed, since the "?" scores were approximately equal on the first administration of the projected and normal testings.

On the L scale, a significant difference exists between the experimental and normal means. The data appear in Table 50. The mean L score obtained under the experimental condition was 3.54 whereas that obtained under the normal condition was 3.16. The difference indicates a systematic bias on the L scale for the paced-projected method with respect to the normal method. This evidence contradicts the traditional claim that the forced first response is the best response. The increase is interpreted as S's tendency to respond positively to the socially desirable items of the L scale when he does not have sufficient time to reflect on the logic of his response.

Table 50

Analysis of Variance of L Scale

Source of Variation	df	Mean Square	F
Sex	1	16.51	2.51
Order	1	3.22	0.49
Sex x Order	1	3.94	0.60
$Error_B$	52	6.58	
Condition	1	3.94	4.37*
Sex x Condition	1	2.01	2.23
Order x Condition	1	0.72	0.80
Sex x Order x Condition	1	0.01	0.01
$Error_W$	52	0.90	

*$p < .05$

The findings of this study suggest that responses on the MMPI will be similar under paced-projected or normal administration conditions for males, but this may not be the case for females. A greater number of omissions and a greater response to L scale items were regarded as providing evidence that normal administration may lead to more valid scores in terms of scores on the MMPI validity scales.

The relationship between changes on the clinical scales noted in this study and the concurrent validity of the scores was not investigated. Further study is necessary to determine the relative effectiveness of scores produced by the two methods of administration in making clinical classifications.

VII. RESPONSE-MARKING TIME

The two studies which appear in this section deal with the establishment of minimum durations for intra-item exposure intervals.

RESPONSE MARKING TIME: TWO CHOICE ITEMS

Problem

The paced-projected method of test administration requires that items either singly or in small groups be exposed to examinees for an interval of time during which they solve the item and indicate their response to it. Setting values for intra-exposure intervals must take into account that portion of the total exposure that theoretically should be devoted to response marking. The study reported here deals with an analysis of how examinees when taking a clerical-speed-and-accuracy type test divide their total test taking time between solving items and marking their responses to items.

Procedure

Subjects. Ss were twenty-one high school pupils of both sexes who were selected to give a normal distribution of mental ability.

Test. Two forms of a clerical speed and accuracy test were constructed. Each form consisted of thirty-five test items and ten practice items. The equivalency of forms was demonstrated in a preliminary study. The items in the test were sufficiently simple to warrant believing that if a high school student attempted any item, he would respond to it incorrectly only because of his carelessness.

Testing conditions. Each S was tested individually and he was administered both the test under control and under projected conditions. The order of the control and projected administrations was counterbalanced over all Ss. Directions for both administrations admonished Ss to work as rapidly and accurately as they could. The control administration was conducted in the usual way except that the test was administered without a time limit. The projected administration consisted of presenting one item at a time and there was no inter-item exposure interval. In advance of the projected administration each S was taught how to use the remote control switch for the slide projector and he was given practice in its use. During the projected administration S had complete control over the duration of the intra-item exposure interval. He triggered the projector to cause an item to be presented and when he responded to it he again triggered the projector to cause the next item to appear, etc. He held the remote control switch in the nonpreferred hand. Operating the control did not seem to impair response marking nor did Ss report that it did when asked directly.

Two Os recorded the solution time and the response-marking time for each item for each S. Their times were averaged and the average is regarded as the obtained measure. Response-marking time was defined to be that interval elapsing between the instant at which S averted his eyes from the screen or test booklet until he completed marking his response.

Solution time was regarded as the interval between completing response marking and subsequently averting his eyes from screen or test booklet. It is recognized that measures of solution time are substantially less valid than are the response-marking times inasmuch as they include time for locating the next item in the test booklet or causing the next item to appear on the screen.

Analysis. Means and standard deviations were computed on response-time and solution-time for each testing condition. Significance tests were computed.

Results

The projected administration required more test-taking time on the average, 115.24 seconds, than did the control administration, 84.03 seconds. The difference is significant at the .01 level. Means, standard deviations, and significance tests appear in Table 51. Significant differences occur between administrations on solution time and response-marking time, respectively, but not on the related variances.

Table 51

Means and Standard Deviations for
Control and Projected Administrations

	Solution Time			Marking Time		
	Control	Projected	P	Control	Experimental	P
Mean	44.50	65.95	.01	40.60	49.29	.01
S. D.	19.90	15.43	NS	8.39	8.21	NS

Over all items, response-marking time on the projected administration ranged from 37.0 to 64.5 seconds. Thus it would seem that on a per-item basis about 1.6 seconds of an intra-item exposure interval should be allotted on two-choice items for response-marking time if one wishes to accommodate the S with the slowest average marking time; however, some Ss on some items took substantially more than 1.6 seconds to mark their responses.

It is interesting to consider that studies reported in earlier sections strongly suggest that projected testing permits covering a given number of items in substantially less time than is required for covering the same items under normal paper-and-pencil conditions without sacrificing validity, reliability, or number of correct responses. Thus it seems that Ss can function effectively at a pace more rapid than the pace they would set as optimal for themselves.

RESPONSE MARKING TIME: FOUR-CHOICE ITEMS

Problem

The problem is the same as that reported in the study above except that the item-type was a four-choice multiple choice instead of a two-choice.

Procedure

Subjects. Ss were twenty-nine students of both sexes from grades 9-12 who were selected to represent a normal distribution of ability.

Test. Two forms of a four-choice multiple choice test, of the clerical speed and accuracy type, were constructed. Each form consisted of thirty items and five practice items. The forms were demonstrated to be equivalent in a preliminary study.

Analysis. Means, standard deviations, and significance tests were computed for response times and solution times on the projected and control administrations.

Results

The mean total test-taking time was not significantly different for the control (199.80 seconds) and projected (206.07) administrations. Mean response-marking time on the projected administration (55.88 seconds) was significantly different from mean response-marking time on the control administration (47.90 seconds). Mean solution time did not differ between administrative conditions: projected (150.19) and control (151.90). The data appear in Table 52.

Table 52

Means and Standard Deviations on Projected and Normal Administrations

	Solution Time			Response Time		
	Projected	Control	P	Projected	Control	P
Mean	150.19	151.90	NS	55.88	47.90	.01
S. D.	51.64	58.12	NS	10.57	9.38	NS

Total response-marking time ranged from 37.5 seconds to 79.5 seconds on the projected administration. Thus, it would seem 2.6 seconds of an intra-item exposure interval might be allotted for response-marking time on four-choice items if one wishes to accommodate the S with the slowest average response time.

VIII. CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Inasmuch as the broad purpose of the project was the exploration of a relatively uncharted area, namely the projected mode of test administration, the studies conducted under the project were primarily exploratory rather than definitive in character. The results of these exploratory studies can only be stated as promising leads, probably blind alleys, and more or less intuitive judgments. This fact was foreseen and the consequences deliberately, but at times reluctantly, accepted by the Es because they believed that initial studies of an exploratory character would be more productive over the long time than would be an exhaustive study of a given condition, the relative importance of which was not known. Other investigators may well have made a different choice. The present investigators hope that their several studies and their statements incorporated in this section may prove useful to future investigators by enabling them to focus more directly upon pertinent variables and to devise productive and efficient studies and applications.

With regard to the equipment which we used, we feel that if we were beginning again, but with benefit of the experience which we had, we would continue to use the Polaroid equipment for slide preparation but would do all projection by means of a remotely controlled automatic 35 mm slide projector coupled to a tape recorder by means of a programmer. Such an arrangement would conserve manpower in presenting the test materials; it would reduce any distraction provided by a two-man operation, and it would assure that intra- and inter-item exposure intervals were of uniform duration over testing sessions. However, this is not to be construed to mean that these studies which incorporated the manually operated projector are deficient due to the equipment which was used. The comment is intended merely to convey our thought that automatic projection equipment which is triggered by a magnetic tape provides great convenience and uniformity of conditions over testing sessions in the case where the latter is needed.

The studies dealing with the presentation of mental ability tests seemed on the whole to indicate that such tests might be adapted readily to the projected mode of test administration. Two troublesome aspects of adapting such tests are (a) presenting computational items (which is discussed subsequently) and (b) adapting power-speed tests for projected administration. With regard to the latter, we feel that if it can be demonstrated that the factor structures of the scores from such a test administered normally and scores from the same test administered normally but with extended time limits are not significantly different from each other, then the adaptation is feasible. We feel this to be true because scores from the normal administration with extended time limits and the sets of scores from various projected administrations did not generally differ from each other.

One general comment about our perception of the interchangeability of the normal and projected testing modes is in order. We would regard one method as interchangeable with the other when the correlation between the sets of scores obtained under the two administrative conditions is high. We regarded "high" as approaching or equal to the equivalence reliability of the test when determined on a test-retest basis, both administrations being executed under normal conditions. Some of our studies demonstrate equivalence of the two testing conditions, according to the definition above, although pairs of means and variances were significantly different. Therefore, it seems to us that new "projected" norms would have to be established for tests in order to permit comparing performances on a given test when administered under different conditions. This point is important but very obvious. Our excuse for the extended presentation of it is simply that we have repeatedly witnessed test users treating raw scores from two forms of a given test as though they were interchangeable when in the raw score state.

The studies dealing with the adaptation of achievement tests were generally gratifying. The studies indicated that the projected-audio presentation seemed not to lead to scores different from the projected-silent or normal administrations. However, this point merits further investigation. It was observed and corroborated by the testees themselves, that some Ss deliberately attended only to one of the stimuli, e.g., sound and not sight or vice versa. This raises the question of whether the audio-visual presentation was, in fact, audio for some, visual for others, and audio-visual for none or for very few. It is our general impression, based on admittedly flimsy evidence, that the more capable Ss attended primarily to the visual stimulus. Some of the former claimed that the audio slowed them; i.e., it generally interfered with their performance perhaps by causing them to pace themselves to the audio which was at a slower rate than their normal reading rates. It is probable that less capable Ss, who might have reading difficulties, would be aided by the audio presentation inasmuch as their spoken comprehension is probably higher than their silent reading comprehension and inasmuch as the audio would permit them to come to grips more quickly with the item-problem. It is interesting to consider the interpretation which might be laid on achievement test scores of students who were benefited, assuming it could be demonstrated, by the audio supplementation of the visual stimulus. Would the obtained scores be a more valid index of achievement than would normally obtained scores because the former is not affected by low reading comprehension which the test does not purport to measure anyway, or would the score be less valid because it is likely that success in most scholastic pursuits is markedly dependent on reading comprehension and ridding it from the test score leads to a misleading appraisal? No doubt resolving this hypothetical problem depends on stating the kind of validity one has in mind; however, it is conceivable that in the imagined case the content validity of the score would be increased and the predictive validity for scholastic criteria would be decreased.

The studies bearing on the adaptation of arithmetic tests to projected administration yielded encouraging results although there is obvious need for more penetrating studies that will determine the best way to adapt computational items for the experimental mode of testing. The technique of revising computational items into recognition-of-correct-computation items was successful in the sense that a purer measure of quantitative ability was obtained; this was achieved, albeit accidentally, at the sacrifice of equivalency of scores from the two sets of measurement, i.e., from the normally administered computation items and the projected recognition items. Certainly another avenue of research in this area centers on the feasibility of adapting to recognition items the verbally presented arithmetic items--as contrasted with the simple presentation of numbers and an operation sign--so that they might be presented by the projected method. Perhaps when items of that kind are presented, the examinee should be requested to indicate if the form or the structure of the proposed solution is correct instead of the correctness of the computations in the solution. This suggestion is based on the obvious differences in kinds of abilities which are evoked by a work problem, on the one hand, and the same arithmetic problem set up as numbers and operation sign, on the other. Parenthetically, it might be added that if projected testing does necessitate substituting recognition-of-correctness items for computational items, then such substitution, despite whatever empirical warrant might be adduced for it, will be greeted coldly. Consider the lack of acceptance of multiple-choice items which appraise writing ability despite the validity studies which give clear mandate for the substitution.

Although no date reported in earlier sections of this report deal with this point, it is appropriate to mention that the American Government study, which was reported earlier, attracted a large number of inquiries from college faculty about whether tests might be administered in that way to their classes. The reasons underlying their inquiries are interesting to consider. First, they reasoned that it would be easy to maintain test security if tests were on slides rather than paper. Second, they reasoned that over a few semesters they could build a sizeable pool of valid and discriminating items from which they could draw when they wished to construct a test. Third, they reasoned that the expense of testing would be reduced over the long run if the projected mode of testing were adopted.

The projected administration of a vocational interest inventory yielded results which were as good or better from both the equivalence and stability standpoints than were data yielded by the administration of the inventory under normal conditions. However, no evidence was gathered about the predictive validities of the obtained measures.

A study in adjustment and personality dealt with a comparison of scores obtained from a delinquency-proneness scale which was administered under normal and projected conditions to normals and delinquents. Another study which was presented was essentially an extension and cross validation of the first study. The data revealed that the scores of normals increased considerably when the test was presented under paced-projected conditions. An inference was made that the paced-projected mode of testing denied Ss time to reflect leisurely about the nature of socially acceptable responses to the items--and as a consequence their impulsive responses tended toward indicating delinquency proneness. This phenomenon is now being investigated under terms of another grant which was awarded the investigators. Apart from the particular significance of the findings to such matters as detection of delinquency proneness and the prediction of future deviate behavior, the finding suggests that the projected mode of testing might enable appraising human attributes which cannot be appraised presently when the usual paper-and-pencil mode of test presentation is employed.

The studies on response and solution times were done to establish "minimal" durations of the intra-item exposure interval for two-choice and four-choice multiple choice items. "Minimal," as used here, meant time-to-mark-a-response without regard to any consideration of the time needed to solve the item. The noteworthy aspect of the results of these studies was that the results were so frequently flaunted by Ss who participated in other studies. For example, the average marking-response time for two-choice items based on Ss who were urged to work as rapidly as they could was about of the same duration or longer than was the threshold time which was noted for a highly complex verbal ability **test** and which included both the solution time and the marking time. In short, the studies, except for the study dealing with vocational interests, revealed that Ss can be paced over items of tests, with no resulting decrement in score, at a far more rapid rate than that at which they work normally and at a more rapid rate than that rate which they set for themselves when requested to work as rapidly as they can.

The studies on response time were done on specially constructed clerical speed and accuracy tests. The focus was response time and not the determination of whether tests of that type could be adapted for projected administration. Nevertheless, it became apparent that such tests could be readily adapted in one of several ways. For example, items in such a test could be split into blocks of items, say five items to each block. Presentation of items might be accomplished in either of two ways. First, blocks could be exposed one-at-a-time for decreasing intervals of time such that on the final block no one could respond to any item which appeared. Or items within a block could be exposed singly and each item in the block would have a constant duration; increasingly less time would be accorded succeeding blocks so that on the last block no S could respond. On the face of it the former method would be preferable because it would sustain motivation for a longer period.

Ignoring the specific studies and the particulars with which they dealt, broad conclusions might be drawn that are more or less generally valid. Several such summary statements are presented below.

The projected mode of test presentation appears to be feasible for the administration of tests of cognitive abilities. When the method is applied to measures of attitudes, the results might or might not be equivalent with normally obtained results and they might or might not be more valid than normally obtained results. This indefiniteness is attributable to the vocational interest and the delinquency-proneness studies which, although yielding gratifying results, cast some doubt on whether scores so obtained were comparable to normally obtained results. As a matter of fact, it is the belief of the Es that the experimental test scores from those tests were superior in many respects to the normally obtained scores.

The projected mode of test administration can yield a greater coverage of test content in a unit of time than can the normal mode of test presentation. It would seem that for cognitive tests, test-taking time can be reduced without sacrifice of mean, variance, reliability, or validity. Further reduction can be made without altering the reliability and validity of scores although their distribution might be altered.

Making intra-item exposure interval a function of item difficulty seems to produce no technical advantage over constant intra-item exposure intervals. However, no conclusion is warranted about the goodness of making intra-item exposure interval a function of item reading load. It is our subjective opinion based on observation that variable intra-item exposure intervals maintains Ss alertness at a higher level than do constant intervals.

Test content and item format are variables which are as important as variables such as intra-item exposure interval and items per exposure when adapting a paper-and-pencil test for projected administration. Ss claim, although without benefit of first-hand experience, that multiple choice items might best be presented by splitting the screen vertically so that the item would appear on the left and the alternatives would appear on the right. They claim this would reduce the time involved in searching for a response, reviewing the alternatives in light of the problem, etc. That test content is an important variable is adequately demonstrated by the arithmetic achievement studies and studies of the quantitative parts of mental ability tests.

The projected mode of administration increases the number of items which Ss attempt. Although this would be expected on the basis of exposing all items to Ss, there seems to be an additional advantage in that the procedure weakens the "cautiousness" response set of some Ss, perhaps because Ss know that an item will be presented one, and only one, time and it prevents persevering on difficult items which appear early in the test.

The exposure of three items simultaneously--one difficult, one of average difficulty, and one simple--appeared to yield results most like normally obtained results and it elicited more favorable comments from Ss than did exposure of single items. This generalization is offered quite hesitantly because it is based on only one study; however, in that study Ss, who had been subjected to both triad and single item exposure, voiced overwhelming favor for the triad exposure. The multiple exposure presentation was discarded, despite the favorableness with which it was received, because the screen presentation of it seemed to us to be difficult to read due to the large number of lines of material which appeared on the screen. This difficulty might have been remedied had we had a larger portable screen at our disposal. The imminence of wall-size TV screens will certainly resolve the problem. Although not having unassailable data to support our opinion, we believe that the triad presentation met with favor because less capable Ss could respond to items through the last triad because a simple item appeared as the first item in each triad. In short, motivation was probably sustained throughout whereas when items were presented singly and in ascending difficulty, some Ss were beyond their depth quite early in the test and remained in that state until the test terminated. A second probable reason is that the triad presentation permitted Ss to divide their time among the three items as they saw fit, thereby permitting them to respond to, say, two items, and to ignore the third completely. On the other hand, had the items been presented singly, they might have been able to complete their solution of only one of the items.

We believe that the evidence which has been gathered supports the above generalizations. However, we are aware that the degree of certainty of a proffered generalization differs from one to the other.

We do have some reservations about procedures which were used and the implications of results which were obtained. Generally the projected tests were of a short duration. We wonder about the extent to which the results we obtained and the sustained S motivation which we observed might be a function of the novelty of the testing procedure rather than a function of some desirable characteristic inherent in it. Further research might profitably be devoted to this point. Perhaps it might be appropriate to administer a battery of tests by the projected method at a single session, as is frequently done under normal administrative conditions, and then to compare the results with those obtained normally and with first-test last-test differences on the projected battery.

We are also aware of the fact that testing by television may precipitate new financial, legal, and ethical problems. If commercial publishers were to package tests for projected administration, and if such tests were broadcast over open-circuit educational TV networks, then one might anticipate a host of problems which would ensue. Would the test materials remain secure? If not, would it be necessary for publishers to prepare new forms annually for projected administration? Would a test so packaged be sold on a per-use basis, or on a number-of-students to be tested basis, or on an outright purchase basis? By the same token, on what bases might a test author and a test publisher agree on the test author's remuneration for devising the test?

Assuming, for the moment, open circuit broadcasts over E-TV networks, one might inquire into the ethical problem which would confront the director of such testing. If one might assume that he would be a member of the American Psychological Association, then the question arises as to whether he would violate the Ethical Standards for Psychologists by divulging test materials indiscriminately; i.e., by open-circuit broadcasting of them.

The remainder of this section is devoted primarily to suggesting lines of investigation which might profitably be pursued as continuations of studies which were described in this report.

Inasmuch as one of our goals was to determine whether tests might beadministered via TV and inasmuch as our studies were done wholly in an analogue situation, a logical next step would be to determine if our results would be corroborated by results obtained when TV is the actual mode of test presentation. We hypothesize that they would. But whether they would be or not, the mode of testing has obvious merit even if confined to the mode of projection which we employed.

Another avenue of investigation centers on beneficial adaptations of the process we employed. Two possible innovations are advanced for consideration. First, a warning device might be included in the mode of presentation so that Ss would be informed of the impending removal of an item from the screen. If test materials are reproduced on moving picture film, then the device might take the form of a time bar marked in seconds which would appear at the top of the item and which would decrease in length in proportion to the remaining duration of the exposure interval. Also, an audible warning device might be used to similar advantage. This suggestion is based on our observations of student behavior during a projected test. When no signal is used, Ss who respond before termination of the exposure sometimes fail to attend to the screen and as a consequence the next item is presented without their being aware of it; when items require scratch-work, the point of attention is the scratch-paper, not the screen, and in some cases Ss continue their scratch-work long after the item has disappeared; and, finally, the warning would further speed Ss into responding. Second, response indication is sometimes difficult depending on the nature of the response and it seems that this difficulty might be reduced. For example, a pre-coded answer sheet would eliminate the need for referring from scratch-work to screen to answer sheet in order to indicate a response, then a "checking" repetition of the cycle--all of which is liable to errors in reading and memory. Another possibility, when scratchwork is unnecessary, would be development of a push-button mode of responding that would imprint the intended response on the answer sheet, thus eliminating the use of pencil and the search for the correct answer space. With such a device, Ss could attend to the screen throughout the test instead of dividing their time between it and the answer sheet.

We feel that individual testing machines are a distinct possibility since some of the variables underlying their use have been investigated in these studies and, additionally, some of them are being studied intensively in the teaching machine context. A simple machine might be built that would receive a test, perhaps printed on a reel as are some teaching machine programs, and would present it item-by-item to an individual student. Such features as duration of intra-item exposure interval, etc., could be set on the machine or the test item itself could adjust the machine for the proper duration. Similarly, scoring could be done by the machine as responses were indicated, e.g., by pushbuttons. Machines of this type would probably be useful to a teacher who wishes supplementary test information of certain kinds from a few students; also, it would be useful to guidance personnel who need additional test information from counselees yet have little time for administration of the tests. It is our feeling the technical features of such a machine are simple and that sufficient information is in hand about the crucial variables related to testing by that mode.

Further investigation might be done to ascertain the unique conditions, i.e., the values to be assigned the underlying variables, under which projected testing (a) can be carried out for a maximum duration at a single testing session and (b) can be reduced to the absolute minimum time to obtain certain information of high reliability and validity. A weakness in our study is the brevity of the testing session to which we confined ourselves. In this context, it might be appropriate to suggest that the inter-item exposure interval, which apparently made no difference in our studies, might become an extremely important variable when the testing session is quite lengthy.

Projected testing opens many avenues for testing which hitherto have been closed on account of the limitations imposed by the usual booklet presentation of test materials. It affords the possibility of animated items. For example, items may take this form: a thirty-second moving picture sequence of a pilot's instrument panel, then a question of "What should he do now, if he wishes to accomplish so and so?" Similarly, lengthy reading passages might be rolled over the screen in order to appraise reading comprehension. Generally, situational tests become a distinct possibility because it would be possible to depict in the testing situation the on-going operation of whatever the case in point, e.g., movement of machinery, the interchanges in a social group, etc. By the same token the addition of sound to such items adds a new dimension to the test situation; e.g., it might be possible to assess whether the prospective pilot can attend to, interpret, and integrate simultaneously information fed to him by sight from his instrument panel and by sound from his earphones.